Chapter 5

Statistics with Imprecise Probabilities—A Short Survey

Thomas Augustin

Abstract This chapter aims at surveying and highlighting in an introductory way some challenges and big opportunities a paradigmatic shift to imprecise probabilities could induce in statistical modelling. Working with an informal understanding of imprecise probabilities, we discuss the concepts of model imprecision and data imprecision as the two main types of imprecision in statistical modelling. Then we provide a short survey of some major developments, methodological questions and applications of imprecise probabilistic models under model imprecision in the context of different inference schools and summarize some recent developments in the area of data imprecision.

5.1 Introduction

By promising powerful solutions to some of the deepest foundational problems of probability and statistics, imprecise probabilities offer great opportunities also for the area of statistical modelling. Still, in statistics, theories of imprecise probabilities live in the shadows, and admittedly the development of many of the imprecise probability-based methods is often in a comparatively early stage. Nevertheless, in all areas of statistics, the desire for a more comprehensive modelling of complex uncertainty had popped up again and again. The rather scattered work covers a huge variety of methods and topics, ranging from contributions to the methodological and philosophical foundations of inference to very concrete questions of applications.

T. Augustin (✉)

Department of Statistics, Ludwig-Maximilians-Universität München (LMU Munich),
Munich, Germany

e-mail: augustin@stat.uni-muenchen.de

© The Author(s) 2022

L. Aslett et al. (eds.), *Uncertainty in Engineering*,

SpringerBriefs in Statistics,

https://doi.org/10.1007/978-3-030-83640-5_5

The present chapter aims at providing a rough and informal survey of some of the major questions and developments.¹ It is structured as follows. Section 5.2 collects the basic concepts that are needed later on. Then, Sect. 5.3 looks at the major sources of imprecision in statistical modelling. We distinguish there between several types of data imprecision and of model imprecision. Section 5.4 focuses on the issue of model imprecision and discusses it from the angle of different inference schools. We put some emphasis on the (generalized) frequentist and Bayesian setting, but also briefly adopt other perspectives. In Sect. 5.5, approaches to handle so-called ontic and epistemic data imprecision, respectively, are surveyed. Section 5.6 is reserved for some concluding remarks.

5.2 Some Elementary Background on Imprecise Probabilities

In this section, we briefly summarize the concepts in the background. With respect to the basic notions and the technical framework for statistical inference, we refer to the first chapter of this book [41]. We rely on the same basic setting, where we use observations (data) on some underlying stochastic phenomenon² to learn characteristics of that mechanism, mathematically described by an unknown parameter of the underlaid probability model.

With respect to imprecise probabilities, a very rough and eclectic understanding shall be sufficient to read this chapter.³ It is not necessary, for our aims here, to distinguish different approaches with respect to many technical details. Thus, in a rather inclusive manner, we subsume here under *imprecise probabilities* any approach that replaces in its modelling precise, traditional probabilities $p(\cdot)$ by *non-empty sets* $\mathcal{P}$ of precise probabilities as the basic modelling entity, including also all approaches that can be equivalently transferred into a set of precise probability. This comprises approaches directly working with sets of probabilities, like robust Bayes analysis (see Sect. 5.4.2), Kofler & Menges' linear partial information (e.g. [42]), Levi's approach to epistemology (e.g. [45]), as well as the whole bunch of corresponding approaches based on non-linear functionals and non-additive set-functions, covering lower and upper previsions in tradition of Walley's book [66], interval probabilities building on

¹ A longer survey discussing the state of the art of statistical inference with imprecise probabilities at that time is aimed at by [8].

² To avoid any commitment to a certain interpretation of probability, we use the term "stochastic phenomenon" as a kind of neutral, superordinate concept, in particular including approaches referring to random phenomena only as well as approaches addressing generally situations of epistemic uncertainty.

³ An introduction into imprecise probabilities on an intermediate level is aimed at by [4]; see also [57] (in this volume) for an introduction, [15] for a survey by a philosopher and [9] for a review with an engineering background. Current developments in research on imprecise probabilities are mostly discussed at the biannual ISIPTA (International Symposium on Imprecise Probabilities: Theories and Applications) meetings; see [1, 5, 27] for the most recent ones.

Weichselberger ([72]⁴), probabilistically interpretable approaches based on capacities including the suitable branch of Dempster-Shafer theory (e.g. [29]), random sets (e.g. [12, Chap. 3]) and p-boxes following [32]. Moreover, there is a smooth transition to several approaches propagating systematic sensitivity analysis (e.g. [52]).

It is important that our basic entity, the set $\mathcal{P}$, has to be understood and treated as an entity of its own. It is by no means possible to distinguish certain of its elements as more likely than others or to mix its elements, eventually leading to a precise traditional probability distribution. $\mathcal{P}$ may be assessed directly by collecting several precise models, so-to-say as possible worlds/expert opinions, to be considered. More often $\mathcal{P}$ is constructed, typically as the set of all probability distributions

- that respect bounds on the probabilities of certain classes of events or, more generally, the expectations of certain random variables,^{5, 6}
- or that are (in a topologically well-defined sense) close to a certain precise probability distribution $p_0(\cdot)$ (*neighbourhood models*), providing a formal framework for expressing statements like “ $p_0(\cdot)$ is approximately true”,⁷
- or that are described by a parameter ϑ varying in an interval/cuboid (*parametrically constructed models*), like “imprecise versions” of a normal distribution.⁸

5.3 Types of Imprecision in Statistical Modelling

There are many situations in statistics where imprecision occurs, i.e. where a careful modelling should go beyond the idealized situation of perfect stochasticity and data observed without any error and in an ideal precision. To study these situations further, an ideal-typical distinction between model and data precision is helpful.

Model Imprecision has to be taken into account whenever there is doubt in a concrete application that the strong requirements (perfect precision and, by the additivity axiom, absolute internal consistency) the traditional concept of probability calls for can be realistically satisfied. This includes all situations of incomplete or conflicting information, robustness concerns, and repeated sampling from a population where the common assumption of i.i.d. repetitions is violated by some unobserved heterogeneity or hidden dependence. In a Bayesian setting, in addition, quite often, if at all, a precise prior distribution is not honestly deducible from the knowledge actually available. Ellsberg’s [31] seminal thought experiments on urns with only partially known compositions have made it crystal clear that the amount of ambiguity, i.e. the

⁴ See also [73] for a summary in English language.

⁵ For instance, in the approaches founded by [66, 72], respectively.

⁶ It may be explicitly noted that the latter also includes comparative/ordinal probabilities (e.g. [49]), only ordering the probability of certain events like “ A is more likely to occur than B , and B more likely than C ”.

⁷ See Sect. 5.4.1.

⁸ See, for instance, the sets of conjugated distributions in Sect. 5.4.2 for an example.

uncertainty about the underlying stochastic mechanism, plays a constitutive role in decision making and thus has to be modelled carefully.

Data Imprecision comprises all situations where the observations are not available in the granularity that is originally intended in the corresponding application. Following [26], for the modelling of data imprecision, it is crucial to distinguish two situations, the precise observation of something inherently imprecise (*ontic data imprecision*) and the imprecise observation of something precise (*epistemic data imprecision*).⁹ The difference between these two may be explained by a pre-election study where some voters are still undecided which single party they will vote for. (Compare [55], and for more details [44, 53]). If we understand a voting preference like “I am still undecided between parties *A* and *B*” as a political position of its own, then we interpret the information as an instance of ontic imprecision. If we focus on the forthcoming election and take this information as an imprecise observation allowing us to predict that the voting decision on the election day will either be *A* or *B*, then we are coping with epistemic imprecision. In particular in engineering, another frequent example of epistemic data imprecision occurs from insufficient measurement precision, where intervals instead of precise values are observed. A contrasting example where (unions of) intervals are to be understood as an entity of their own arises when the time span of certain spells is characterized: “This machine was under full load from November 10th to December 23rd.”

Epistemic imprecision very naturally occurs in many studies on dynamic processes, from socio-economic over medical to technical studies. There, censoring is always a big issue: the spells of some units are still unfinished when the study ends, providing only lower bounds on the spell duration. Typical examples include the duration of unemployment, the time to recurrence of a tumour or the lifetime of an electronic component. In addition, also interval censoring is quite common, where one only learns that the event of interest occurred within a certain time span. It should be noted explicitly that missing data, a frequent problem for instance in almost every social survey, may be comprised under this setting, taking the whole sample space as the observation for those units missing.

5.4 Statistical Modelling Under Model Imprecision

Of course, there are also strong reservations against imprecise probabilities in traditional statistics. For a traditional statistician, imprecise probabilities are just a superfluous complication, misunderstanding either the generality of the concept of uncertainty or the reductionist essence of the modelling/abstraction process. Indeed, for an orthodox Bayesian, all kinds of not-knowing are simply situations under uncertainty, and any kind of uncertainty is *eo ipso* expressible by traditional probabilities. From the modelling perspective, imprecision is taken as part of the residual category

⁹ See also the distinction between the disjunctive and conjunctive interpretation of random sets, as discussed, e.g. in [30, Sect. 1.4].

that is naturally lost when abstracting and building models.¹⁰ Box (and Draper)’s often cited dictum “Essentially, all models are wrong, but some of them are useful” [14, p. 424] has generally been (mis)understood as a justification to base details of the model choice on mathematical convenience and as an irrevocable argument to take model imprecision as negligible.

5.4.1 *Probabilistic Assumptions on the Sampling Model Matter: Frequentist Statistics and Imprecise Probabilities*

Indeed, however, Box’s quotation could be continued by “...and some models are dangerous”, since the general neglecting of model imprecision implicitly presupposes a continuity (in an informal sense) of the conclusions in the models that by no means can be taken as granted. A well-known example from robust statistics¹¹ is statistical inference from a “regular bell-shaped distribution”. The standard proceeding would be to assume a normal distribution. But, for instance, the density of a Cauchy distribution is phenomenologically de facto almost indistinguishable from the density function of a normal distribution, suggesting both models to be equivalent from a practical point of view. However, statistical inference based on the sample mean shows fundamentally different behaviours.¹² Under normality, the distribution of the sample mean behaves nicely, contracting itself around the correct value if the sample size n increases. In the case of the Cauchy distribution, however, the distribution of the sample mean stays the same irrespective of the sample size,¹³ making any learning by the sample mean impossible.

This shocking insight—optimal statistical procedures may behave disastrously even under “tiny deviations” from the ideal model—demonstrates that imprecision in the underlying model may matter substantially. In this context, the theory of (frequentist) robust statistics as the theory of approximately true models emerged, and imprecise probabilities provide a natural superstructure upon it (see, e.g. [38] for historical connections). In particular, neighbourhood models, also briefly mentioned above, have become attractive.¹⁴ Building on an influential result by Huber and Strassen [39], a comprehensive theory of testing in situations where the hypotheses

¹⁰ This is particular the case when the prescriptive character of models is emphasized over its descriptive role (cf., e.g. [13, p. 99]).

¹¹ See also [8, Sect. 7.5.1, in particular, Fig. 7.3] to illustrate this.

¹² Generally, if one looks at an i.i.d. sample $X_1, \dots, X_n$ from a normal distribution with parameters μ and σ^2 , denoted by $\mathcal{N}(\mu, \sigma^2)$, and an i.i.d. sample $Z_1, \dots, Z_n$ from a Cauchy distribution with parameter α and β , denoted by $\mathcal{C}(\alpha, \beta)$, respectively, one obtains for the sample means $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ and $\bar{Z} := \frac{1}{n} \sum_{i=1}^n Z_i \sim \mathcal{C}(\alpha, \beta)$.

¹³ The Law of Large Numbers is not applicable to the Cauchy distribution because it does not have moments.

¹⁴ See, e.g. [6] for a survey.

are described by imprecise probabilities emerged (see [2], [8, Sect. 7.5.2] and the references in the corresponding review sections therein.) Further insights are provided from interpreting frequentist statistics as a decision problem under an imprecise sampling distribution, leading to the framework investigated in [34].

Other frequentist approaches, starting from different angles, include Hampel's frequentist betting approach (e.g. [36]) and some work on minimum distance estimation under imprecise sampling models (e.g. [35]). The statistical consequences of the chaotic models in the genuine frequentist framework to imprecise probabilities, developed by Fine and followers (e.g. [33]), might be quite intriguing, but are still almost entirely unexplored.

5.4.2 *Model Imprecision and Generalized Bayesian Inference*

Priors and Sets of Priors, Generalized Bayes Rule. The centrepiece of Bayesian inference is the prior distribution. Apart from very large sample sizes, where the posterior is de facto determined by the sample, the prior naturally has a strong influence on the posterior and on all conclusions drawn from it. In the rare situations where very strong prior knowledge is available, it can be used actively, but most often the strong dependence on the prior has been intensively debated and criticized.

Working with sets Π of prior probabilities (or interval-valued priors) opens new avenues here. This set can naturally be chosen to reflect the quality/determinacy of prior knowledge: strong prior knowledge leads to “small” sets; weak prior knowledge to “large” sets. Typical model classes include neighbourhood models or sets of parametric distributions which often are conjugate¹⁵ to the sampling distribution, which typically still is assumed to be precise.¹⁶ In imprecise probability, Π is understood as naturally inducing the set Π_x of all posteriors arising from a prior in Π .¹⁷ Interpretations of Π and Π_x vary to the extent they are understood as principled entities. A pragmatic point of view sees an investigation of Π_x just as a self-evident way to perform a sensitivity analysis. On the other extreme, Walley's [66] generalized theory of coherence, having initiated the most vivid branch of research on imprecise probabilities, provides a rigorous justification of exactly this way to proceed as the “Generalized Bayes Rule (GBR)”. Important developments have also been achieved for a variety of different model classes under the term “Robust Bayesian Analysis”; see, e.g. [58] for a review on this topic.

Near Ignorance Models. One important way to use sets of priors is that they allow for quite a natural formulation of (rather) complete ignorance. A traditional Bayesian

¹⁵ See, e.g. [41, Chap. 1.2] (in this volume).

¹⁶ Imprecise sampling models are quite rarely studied, see, however, [66, Sect. 8.5] and [61].

¹⁷ There are some approaches in the traditional Bayesian context that work with sets of priors, too, but merely understanding such a set as an interim step, based on which an ultimate precise prior is selected. This can be done in a data-driven way, as in empirical Bayes approaches, including the ML-II approach (e.g. [11, Sect. 3.5.4]), or generally, as in maximum entropy approaches (e.g. [11, Sect. 3.4]).

model *eo ipso* fails in expressing ignorance/non-informativeness. Assigning a precise probability is never non-committal; every precise prior delivers probabilistic information about the parameter. The genuinely non-informative model is the set of all probability distributions. While this model would yield vacuous inferences, it motivates so-called near-ignorance models, where, informally spoken, the inner core of this set is used, excluding extreme probabilities that are immune to learning. Near-ignorance models still assign non-committal probabilities to standard events in the parameter space, but allow for learning. By far the most popular model is the Imprecise Dirichlet Model (IDM) [67] for categorical data. Different extensions followed, including general near-ignorance models for exponential families (e.g. [10]). Another direction of enabling the formulation of near-ignorance uses all priors with bounded derivatives ([68]; for a general exposition of the concept of bounded influence, see the book [69]).

Prior-Data Conflict. In some sense, a complementary application of generalized Bayesian inference is the active modelling of prior-data conflict. In practice, generalized Bayesian models are quite powerful in expressing substantial prior knowledge. In particular, in areas where data are scarce, it is important to use explicitly all prior knowledge available, for instance by borrowing strength from similar experiments. Then, however, it is crucial to have some kind of alert system warning the analyst when the prior knowledge appears doubtful in the light of data. Indeed, sets of priors can be designed to react naturally to potential prior-data conflict: If data and prior assumptions are in agreement, the set of posterior distributions contracts more and more with increasing sample size. In contrast, in the case of prior-data conflict and intermediate sample size, the set of posterior distributions is inflated substantially, perfectly indicating that one should refrain from most decisions before having gathered further information.¹⁸

5.4.3 *Some Other Approaches*

With generalized Bayesian approaches, and less pronounced with generalized frequentist statistics, the major statistical inference school are also predominant in the area of imprecise probabilities. Nevertheless, there has also been considerable success in other inference frameworks. Again and again, the desire to save Fisher's fiducial argument, aiming at providing probability statements on parameters without having to rely on a prior distribution, has been a driving force for developments in imprecise probabilities. Dempster's concept of multivalued mappings (e.g. [28]), which become even more famous in artificial intelligence by Shafer's reinterpretation founding Dempster-Shafer Theory (see, for instance, again the survey by [29]), is to be mentioned here, but also work by Hampel (e.g. [36]) and by Weichselberger (e.g.

¹⁸ See [71] for a general treatment in one-parameter exponential families, also providing some illustrating plots, and, for instance, [70] for an application in the context of system reliability.

[74]), see also [7] tracing back its roots.¹⁹ A generalized likelihood-based framework has been introduced by Cattaneo (e.g. [16, 17]).

Another direct inference approach is *Nonparametric Predictive Inference (NPI)*, as introduced by Coolen originally for Bernoulli data [18]. Based on exchangeability arguments, NPI yields direct conditional probabilities of further real-valued random quantities, relying on the low structure assumption that all elements of the natural partition produced by the already observed data are equally likely; see, for instance, [19] for a detailed discussion, [3] for a clear embedding into imprecise probabilities and [21] for a web-page documenting research within this framework. The basic approach can be naturally extended to censored data/competing risks (e.g. [22]), and to categorical data [20]. NPI has been developed further in a huge variety of fields; see, for instance, [23, 24] for recent applications in biometrics and finance, respectively.

5.5 Statistical Modelling Under Data Imprecision

In this section, we turn to statistical modelling under data imprecision. Keeping the distinction from Sect. 5.3, we briefly discuss ontic data imprecision and then turn to epistemic data imprecision.

Ontic data imprecision, where we understand the imprecise observation as an entity of its own, may be argued to be a border case between classical statistics and its extensions. Technically, we change the sample space of each observation to (an appropriate subset of) the power set. For instance, recalling the election example from Sect. 5.3, this means that instead of $\{a, b, c, \dots\}$ representing the vote for a single party, we now also allow for combinations $\{a, b\}, \{b, c\}, \dots, \{a, b, c\}, \dots$, representing the indecision between several parties. As long as we are in a multinomial setting, nothing has changed from an abstract point of view, providing powerful opportunities for complex statistical modelling. In the spirit of this idea, [44, 55] apply multinomial regression models, classification trees, regularized discrete choice models from election research, and spectral clustering methods to German pre-election survey data. The situation changes substantially when ordinal or continuous data are considered, because, after changing to the power set, the underlying ordering structure is only partially preserved.

Epistemic data imprecision is, as the examples at the end of Sect. 5.3 show, of great importance in many applications and is quite vividly addressed in classical statistics. Even here, traditional statistics keeps its focus on full identification, i.e. the selection of one single probability model fitting the observed data optimally. One searches for, and then implicitly relies on, conditions under which one gets hands on the so-to-say deficiency process as a thought pattern, making ideal precise observations imprecise. For that purpose, most classical approaches assume either some

¹⁹It may also be argued that, although less explicitly aiming at fiducial reasoning, the work on near-ignorance models discussed above fits well into this category.

kind of uninformativeness of the deficiency process (independent censorship, coarsening at random (CAR) or missingness (completely) at random (MCAR, MAR)) or an explicit modelling of the deficiency process; see the classical work by [37, 46]. Both the uninformativeness as well as the existence of a precisely specifiable deficiency process are very strong assumptions. They are—eo ipso by making explicit statements about unobservable processes—typically not empirically testable. Whenever these assumptions are just made for purely formal reasons, the price to pay for the seemingly precise result of the estimation process is high. In terms of Manski's *Law of Decreasing Credibility*,²⁰ results may suffer severely from a loss of their credibility, and thus of their practical relevance.

Against this background, in almost any area of application, the desire for less committal, cautious handling of epistemic data imprecision arose. Mostly isolated approaches were proposed that explicitly try to take all possible worlds into account in a reliable way, aiming at the *set* of all models optimally compatible with potentially true data. These approaches include, for instance, work from reliable computing and interval analysis in engineering, like [51], extensions of generalized Bayesian inference (e.g. [75]) to reliable statistics in social sciences (e.g. [56]); see also [8, Sect. 7.8.2], who try to characterize and unify these approaches by the concept of cautious data completion, and the concept of collection regions in [60].

There is a smooth transition to approaches that explicitly introduce cautious modelling into the construction of estimation procedures; see, for instance, for recent different likelihood- and loss minimization-based approaches addressing epistemic data imprecision, [25, 40, 43, 54]. Such approaches have the important advantage that their construction often also allows the incorporation of additional well-supported subject matter knowledge, too imprecise to be useful for the precision focused methods from traditional statistics, but very valuable to reduce the set of compatible models by a considerable extent.

Congenial is work in the field of partial identification and systematic sensitivity analysis, providing methodology for handling observationally equivalent models; see [48, 65], respectively, for classical work and [47, 62] for introductory surveys. The framework of partial identification is currently receiving considerable attention in econometrics, where in particular the embedding of fundamental questions into the framework of random sets is of particular importance [50].

5.6 Concluding Remarks

The contribution provided a—necessarily painfully selective—survey of some developments of statistical modelling with imprecise probabilities (in a wider sense, also including closely related concepts). Both in the area of model imprecision as well as under data imprecision, imprecise probabilities prove to be powerful and particularly

²⁰ “The credibility of inferences decreases with the strength of the assumptions maintained.” [48, p. 1].

promising. Further developments urgently needed include a proper methodology for simulations with imprecise probabilities (see [64] for recent results), a careful study of the statistical consequences of the rather far developed probabilistic side of the theory of stochastic processes with imprecise probabilities (e.g. [63]), a more fruitful exchange with recent research on uncertainty quantification in engineering (see, e.g. [59] (in this volume)), an open mind towards recent developments in machine learning and more large scale applications. Not only for these topics it is important to complement the still recognizable focus on so-to-say defensive modelling by a more active modelling. Far beyond sensitivity and robustness aspects, imprecision can actively be used as a strong modelling tool. The proper handling of prior-data conflict and the successful incorporation of substantive matter knowledge in statistical analysis under data imprecision are powerful examples of going in this direction.

References

1. A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors. *ISIPTA '17*, volume 62 of *Proceedings of Machine Learning Research*. PMLR, 2017.
2. T. Augustin. Neyman-Pearson testing under interval probability by globally least favorable pairs: Reviewing Huber-Strassen theory and extending it to general interval probability. *Journal of Statistical Planning and Inference*, 105:149–173, 2002.
3. T. Augustin and F. Coolen. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124:251–272, 2004.
4. T. Augustin, F. P. A. Coolen, G. de Cooman, and M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley, Chichester, 2014.
5. T. Augustin, S. Doria, E. Miranda, and E. Quaeghebeur, editors. *ISIPTA '15*. SIPTA, 2015.
6. T. Augustin and R. Hable. On the impact of robust statistics on imprecise probability models: A review. *Structural Safety*, 32:358–365, 2010.
7. T. Augustin and Rudolf Seising. Kurt Weichselberger’s contribution to imprecise probabilities. *International Journal of Approximate Reasoning*, 98:132–145, 2018.
8. T. Augustin, G. Walter, and F. Coolen. Statistical inference. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, 2014.
9. M. Beer, S. Ferson, and V. Kreinovich. Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing*, 37:4–29, 2013.
10. A. Benavoli and M. Zaffalon. Prior near ignorance for inferences in the k-parameter exponential family. *Statistics*, 49:1104–1140, 2014.
11. J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 2nd edition, 1985.
12. A. Bernardini and F. Tonon. *Bounding Uncertainty in Civil Engineering – Theoretical Background*. Springer, Berlin, 2010.
13. J. M. Bernardo and A. Smith. *Bayesian Theory*. Wiley, Chichester, 2000.
14. G. Box and N. Draper. *Empirical Model-building and Response Surface*. Wiley, New York, 1987.
15. S. Bradley. Imprecise probabilities. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*. Stanford University, 2019. <https://plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities/>.
16. M. Cattaneo. *Statistical Decisions Based Directly on the Likelihood Function*. PhD thesis, ETH Zurich, 2007.
17. M. Cattaneo. Likelihood decision functions. *Electronic Journal of Statistics*, 7:2924–2946, 2013.

18. F. Coolen. Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, 36:349–357, 1998.
19. F. Coolen. On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, 15:21–47, 2006.
20. F. Coolen and T. Augustin. A nonparametric predictive alternative to the imprecise Dirichlet model: The case of a known number of categories. *International Journal of Approximate Reasoning*, 50:217–230, 2009.
21. F. Coolen, P. Coolen-Schrijner, and T. Coolen-Maturi. Nonparametric Predictive Inference (npi), 2020. <https://npi-statistics.com/>.
22. T. Coolen-Maturi. Nonparametric predictive pairwise comparison with competing risks. *Reliability Engineering & System Safety*, 132:146–153, 2014.
23. T. Coolen-Maturi. Predictive inference for best linear combination of biomarkers subject to limits of detection. *Statistics in Medicine*, 36:2844–2874, 2017.
24. T. Coolen-Maturi and F. Coolen. Nonparametric predictive inference for the validation of credit rating systems. *Journal of the Royal Statistical Society: Series A*, 182:1189–1204, 2019.
25. I. Couso and D. Dubois. A general framework for maximizing likelihood under incomplete data. *International Journal of Approximate Reasoning*, 93:238–260, 2018.
26. I. Couso, D. Dubois, and L. Sánchez. *Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables*. Springer, 2014.
27. J. De Bock, C. de Campos, E. de Cooman, G. Quaeghebeur, and G. Wheeler, editors. *ISIPTA '19*, volume 103 of *Proceedings of Machine Learning Research*. PMLR, 2019.
28. A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38:325–339, 1967.
29. T. Denoeux. 40 years of Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 79:1–6, 2016.
30. D. Dubois. Belief structure, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence*, 5:403–416, 1986.
31. D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.
32. S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures (Sandia report), 2003. <https://doi.org/10.2172/809606>.
33. P. Fierens. An extension of chaotic probability models to real-valued variables. *International Journal of Approximate Reasoning*, 50:627–641, 2009.
34. R. Hable. *Data-based Decisions under Complex Uncertainty*. PhD thesis, LMU Munich, 2009.
35. R. Hable. Minimum distance estimation in imprecise probability models. *Journal of Statistical Planning and Inference*, 140:461–479, 2010.
36. F. Hampel. The proper fiducial argument. In R. Ahlswede, L. Bäumler, N. Cai, H. Aydinian, V. Blinovskiy, C. Deppe, and H. Mashurian, editors, *General Theory of Information Transfer and Combinatorics*, volume 4123, pages 512–526. Springer, 2006.
37. D. Heitjan and D. Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19:2244–2253, 1991.
38. P. Huber. The use of Choquet capacities in statistics. *Proceedings of the 39th Session of the International Statistical Institute*, 45:181–191, 1973.
39. P. Huber and V. Strassen. Minimax tests and the Neyman-Pearson lemma for capacities. *The Annals of Statistics*, 1:251–263, 1973.
40. E. Hüllermeier, S. Destercke, and I. Couso. Learning from imprecise data: Adjustments of optimistic and pessimistic variants. In N. Ben Amor, B. Quost, and M. Theobald, editors, *Scalable Uncertainty Management*, pages 266–279. Springer, 2019.
41. G. Karagiannis. Introduction to Bayesian statistical inference. In L. Aslett, F. Coolen, and J. De Bock, editors, *Uncertainty in Engineering – Introduction to Methods and Applications*, pages 1–14. Springer, 2022.
42. E. Kofler and G. Menges. *Entscheiden bei unvollständiger Information*. Springer, Berlin, 1976.
43. D. Kreiss and T. Augustin. Undecided voters as set-valued information – towards forecasts under epistemic imprecision. In K. Tabia and J. Davis, editors, *Scalable Uncertainty Management*. Springer LNCS, 242–250, 2020.

44. D. Kreiss, M. Nalenz, and T. Augustin. Undecided voters as set-valued information - machine learning approaches under complex uncertainty. In Destercke S. and E. Hüllermeier, editors, *ECML/PKDD 2020 Workshop on Uncertainty in Machine Learning*, 2020. <https://sites.google.com/view/wuml-2020/program>.
45. I. Levi. *The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, Cambridge, 1980.
46. R. Little and D. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2nd edition, 2014.
47. C. Manski. Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics*, 191:293–301, 2015.
48. Charles Manski. *Partial identification of probability distributions*. Springer, 2003.
49. E. Miranda and S. Destercke. Extreme points of the credal sets generated by comparative probabilities. *Journal of Mathematical Psychology*, 64:44–57, 2015.
50. I. Molchanov and F. Molinari. *Random Sets in Econometrics*. Econometric Society Monographs. Cambridge University Press, 2018.
51. H. Nguyen, V. Kreinovich, B. Wu, and G. Xiang. *Computing Statistics under Interval and Fuzzy Uncertainty: Applications to Computer Science and Engineering*. Springer, Berlin, 2011.
52. M. Oberguggenberger, J. King, and B. Schmelzer. Classical and imprecise probability methods for sensitivity analysis in engineering: A case study. *International Journal of Approximate Reasoning*, 50:680–693, 2009.
53. J. Plass. *Statistical modelling of categorical data under ontic and epistemic imprecision*. PhD thesis, LMU Munich, 2018.
54. J. Plass, M. Cattaneo, T. Augustin, G. Schollmeyer, and C. Heumann. Reliable inference in categorical regression analysis for non-randomly coarsened observations. *International Statistical Review*, 87:580–603, 2019.
55. J. Plass, P. Fink, N. Schöning, and T. Augustin. Statistical modelling in surveys without neglecting ‘The undecided’. In T. Augustin, S. Doria, M. Miranda, and E. Quaeghebeur, editors, *ISIPTA '15*, pages 257–266. SIPTA, 2015.
56. U. Pötter. *Statistical Models of Incomplete Data and Their Use in Social Sciences*. Habilitation Thesis, Ruhr-Universität Bochum, 2008.
57. E. Quaeghebeur. Introduction to the theory of imprecise probability. In L. Aslett, F. Coolen, and J. De Bock, editors, *Uncertainty in Engineering – Introduction to Methods and Applications*, pages 37–50. Springer, 2022.
58. F. Ruggeri, D. Ríos Insua, and J. Martín. Robust Bayesian analysis. In D. Dey and C. Rao, editors, *Handbook of Statistics. Bayesian Thinking: Modeling and Computation*, volume 25, pages 623–667. Elsevier, 2005.
59. S. Bi and M. Beer. Overview of stochastic model updating in aerospace application under uncertainty treatment. In L. Aslett, F. Coolen, and J. De Bock, editors, *Uncertainty in Engineering – Introduction to Methods and Applications*, pages 115–130. Springer, 2022.
60. G. Schollmeyer and T. Augustin. Statistical modeling under partial identification: Distinguishing three types of identification regions in regression analysis with interval data. *International Journal of Approximate Reasoning*, 56, Part B:224–248, 2015.
61. N. Shyamalkumar. Likelihood robustness. In D. Rios Insua and F. Ruggeri, editors, *Robust Bayesian Analysis*, pages 127–143. Springer, New York, 2000.
62. E. Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2:167–195, 2010.
63. N. T’Joens, J. De Bock, and G. de Cooman. In search of a global belief model for discrete-time uncertain processes. In J. De Bock, C. de Campos, G. de Cooman, E. Quaeghebeur, and G. Wheeler, editors, *ISIPTA '19*, volume 103 of *Proceedings of Machine Learning Research*, pages 377–385, Thagaste, Ghent, Belgium, 2019. PMLR.
64. M. Troffaes. Imprecise Monte Carlo simulation and iterative importance sampling for the estimation of lower previsions. *International Journal of Approximate Reasoning*, 101:31–48, 2018.
65. S. Vansteelandt, E. Goetghebeur, M. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16:953–979, 2006.

66. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London, 1991.
67. P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:3–57, 1996.
68. P. Walley. A bounded derivative model for prior ignorance about a real-valued parameter. *Scandinavian Journal of Statistics*, 24:463–483, 1997.
69. P. Walley. *BI Statistical Methods: Volume I*. Lightning Source, 2015.
70. G. Walter, L. Aslett, and F. Coolen. Bayesian nonparametric system reliability using sets of priors. *International Journal of Approximate Reasoning*, 80:67–88, 2017.
71. G. Walter and T. Augustin. Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 3:255–271, 2009.
72. K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica, 2001.
73. K. Weichselberger. The theory of interval probability as a unifying model for uncertainty. *International Journal of Approximate Reasoning*, 24:149–170, 2000.
74. K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung II. Symmetrische Wahrscheinlichkeitstheorie*. (Manuscript), 2013.
75. M. Zaffalon and E. Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34:757–821, 2009.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.