Computing with sets of probability measures

Fabio Gagliardi Cozman
Escola Politécnica, Univ. de São Paulo
São Paulo, SP - Brazil

José Carlos Ferreira da Rocha
Univ. Estadual de Ponta Grossa
Ponta Grossa, PR - Brazil

Cassio Polpo de Campos
Pontifícia Universidade Católica
São Paulo, SP - Brazil

The computational manipulation of probability measures often requires the treatment of interval values, not only due to numerical errors, but also due to more fundamental difficulties: we may want to model imprecise beliefs; we may have incomplete knowledge about probability values; we may be interested in merging beliefs from groups of experts; and we may wish to verify the effect of perturbations in probabilistic models [2, 17]. Such difficulties have often led to the study of interval probability and related theories. The goal of this paper is to present a brief overview of methods and results that can be relevant to the validated manipulation of probabilistic models.

The most general representation for imprecision in probabilistic models seems to be provided by the theory of sets of probabilities (called *credal sets* [13]). In this work we focus on closed convex credal sets; there are axiomatic derivations of such credal sets and other variants [12, 16, 17]. Consider two examples. First, consider a binary variable X and the set of measures defined by the interval $P(X = x_0) \in [0.3, 0.4]$, where $P(X = x_0)$ is the probability of the event $\{X = x_0\}$ — here a single interval can define the entire credal set. Second, consider a variable Y that can take three values, $\{y_0, y_1, y_2\}$. A probability distribution for Y is entirely defined by a three-valued vector $\{p_0, p_1, p_2\}$ such that $p_i \geq 0$ and $\sum_i p_i = 1$. We can build a credal set by taking a distribution $p(Y)$ and considering the set of all distributions $r(Y)$ such that the difference $|R(A) - P(A)|$ is always smaller than some positive ϵ for any event A (where $R(\cdot)$ is the measure induced by $r(Y)$ and $P(\cdot)$ is the measure induced by $p(Y)$). This type of credal set is called a *total variation* neighborhood in robust statistics [10].

Given a credal set $Q(X)$, we can obtain *upper expectations* for any bounded function: $\overline{E}[f(X)] = \max_{P \in Q} E_P[f(X)]$. Likewise, we can define *lower expec-*

tations: $\underline{E}[f(X)] = \min_{P \in Q} E_P[f(X)]$. Lower and upper expectations define expectations intervals, and the theory of credal sets can be viewed as a theory that manipulates expectation intervals in a principled manner. We assume discrete models in this paper, noting that an assessment of the form $\overline{E}[f(X)] = \gamma$ is equivalent to a linear inequality $\sum_X f(x)p(x) \leq \gamma$.

Conditioning is generally taken to mean elementwise application of Bayes rule; the *conditional* credal set $Q(X|Y)$ is obtained by applying Bayes rule to each element of the *joint* credal set $Q(X, Y)$ [8, 13].

Consider first the computation of upper expectations $\overline{E}[f(X)]$ with respect to credal sets specified by linear constraints. We obtain a linear program with analysis going back to the work of Boole and with extensions based on column-generated methods, as reviewed by Hansen et al [9].

A more interesting challenge is the computation of upper posterior expectations $\overline{E}[f(X)|Y]$. Still assuming linear constraints, we now have a linear fractional optimization problem [15]. The most efficient method to deal with these problems seems to be the Charnes–Cooper transformation, which reduces the fractional problem to a linear program [11, 14]. Other methods, such as Walley's iterative scheme and the Dinkelbach-Jagannatham algorithm (known in statistics as Lavine's method) can be of value in specific cases [6].

An important situation in practice is the computation of $\overline{E}[f(X)|Y]$ with respect to a credal set $Q(X)$ and a collection of "likelihood" credal sets $Q(Y|X = x)$, for all values of X. Surprisingly, we can still reduce this problem to a linear program with some mild assumptions on the sets $Q(Y|X = x)$, using an algorithm presented in [6].

We now consider the impact of independence relations. The first difficulty is that there are several definitions of independence for credal sets [5].

One possible definition (*epistemic independence*) states that variables X and Y are independent when $\overline{E}[f(X)|Y] = \overline{E}[f(X)]$ and $\overline{E}[g(Y)|X] = \overline{E}[g(Y)]$ for any bounded functions $f(X)$ and $g(Y)$. Algorithms for inference in multivariate models based on epistemic independence are presented in [7], but their computational complexity seems to be quite high. A simple Markov chain as $W \to X \to Y \to Z$, where all variables are binary, all probabilities are defined by intervals, and each variable is epistemically independent of all ascendants given the direct ascendant, defines a credal set $Q(W, X, Y, Z)$ with more than 6 million extreme points!

A second possible definition for independence (*strong independence*) requires that any extreme point of $Q(X, Y)$ satisfies $p(X|Y) = p(X)$ and $p(Y|X) = p(Y)$. Computation of upper posterior expectations is now a multilinear program with many possible local maxima. There has been great effort to solve such programs when multivariate models are represented by directed graphs (following the successful theory of Bayesian networks). Exhaustive algorithms have been implemented; the JavaBayes system, freely distributed by the first author at http://www.cs.cmu.edu/~javabayes, offers some support for strong independence. Simulated annealing and genetic search have also been tested [3, 4]. Although the optimization problem is a reverse geometric program [1], geometric duality cannot be easily used here, because the number of dual variables is potentially huge. The most promising approach seems to be branch-and-bound algorithms, coupled with redundancy-elimination computations. Because graphical structures can be used to generate bounds on probabilities, it is possible to gradually "cut" the sizes of credal sets when looking for a global maximum. At the same time, convex hull algorithms can be used to eliminate redundant vertices from credal sets.

References

[1] M. Avriel. *Advances in Geometric Programming.* Plenum Press, New York, 1980.

[2] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 1985.

[3] A. Cano, J. E. Cano, and S. Moral. Convex sets of probabilities propagation by simulated annealing. *Proc. of the Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 4–8, 1994.

[4] A. Cano and S. Moral. A genetic algorithm to approximate convex sets of probabilities. *Proc. of the Int. Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2:859–864, 1996.

[5] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. *Proc. of the First Int. Symp. on Imprecise Probabilities and Their Applications*, pages 121–130, Ghent, Belgium, 1999.

[6] F. G. Cozman. Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics*, 8(4):824–838, 1999.

[7] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[8] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? *Bayesian Statistics*, pages 17–38. University Press, 1980.

[9] P. Hansen, B. Jaumard, M. P. de Aragão, F. Chauny, and S. Perron. Probabilistic satisfiability with imprecise probabilities. *International Journal of Approximate Reasoning*, 24(2-3):171–189, 2000.

[10] P. J. Huber. *Robust Statistics.* Wiley, New York, 1980.

[11] B. Jaumard, P. Hansen, and M. P. de Aragão. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3(2):135–148, Spring 1991.

[12] V. P. Kuznetsov. *Interval Statistical Methods.* Radio i Svyaz Publ., (in Russian), 1991.

[13] I. Levi. *The Enterprise of Knowledge.* MIT Press, Cambridge, Mass., 1980.

[14] M. P. Pacifico, G. Salinetti, and L. Tardella. Fractional optimization in Bayesian robustness. Technical Report Serie A n. 23, Dipartamento di Statistica, Probabilita e Statistiche Applicate, Universita di Roma La Sapienza, Italy, 1994.

[15] S. I. Schaible and W. T. Ziemba. *Generalized Concavity in Optimization and Economics.* Academic Press, 1981.

[16] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23(6):2168–2217, 1995.

[17] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.