Algorithms for Imprecise Probability
Part I

Fabio G. Cozman - Cassio P. de Campos
fgcozman@usp.br - cassiopc@gmail.com
Overview

- Part I: algorithms without independence (this talk).
- Part II: algorithms with independence (next talk, by Cassio).
Overview (some more)

Part I: algorithms without independence (this talk).
1. The basic linear fractional program.
2. Dealing with probabilities that may be zero.
3. Special important cases: neighborhoods, capacities, and the like.
4. Decision making.

Part II: algorithms with independence (next talk, by Cassio).
Easy warm-up

- Possibility space $\Omega$ with states $\omega$; events are subsets of $\Omega$.

- Random variables and indicator functions.
  - Bounded function $X : \Omega \rightarrow \mathbb{R}$.
  - Special type: indicator function of event $A$:
    - Denoted by $A$ as well.
    - $A(\omega) = 1$ if $\omega \in A$; $0$ otherwise.
Axioms for expectations

**EU1** If $\alpha \leq X \leq \beta$, then $\alpha \leq E[X] \leq \beta$.


Some consequences:

1. $X \geq Y \Rightarrow E[X] \geq E[Y]$.
2. $E[\alpha X] = \alpha X$. 
Probabilities

- The probability $P(A)$ is $E[A]$.

Properties of a probability measure:

- **PU1** $P(A) \geq 0$.
- **PU2** $P(\Omega) = 1$.
- **PU3** If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$. 
Conditional expectations/probabilities

- Conditional expectation of $X$ given $B$,

$$E[X|B] = \frac{E[BX]}{P(B)} \quad \text{if} \quad P(B) > 0.$$ 

- Bayes rule: If $P(B) > 0$, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$
Propositional formula $\phi$:
1. propositions
2. operators ($\neg$, $\land$, $\lor$, $\rightarrow$).

Take $\Omega$ as the set of $2^n$ truth assignments for $n$ propositions.

Interpret $P(\phi) \geq \alpha$ as

$$\sum_{\omega \models \phi} P(\omega) \geq \alpha.$$
Probabilistic satisfiability

- Given $m$ assessments, is there a probability measure over $\Omega$?
- Each assessment is a linear constraint.
- Must satisfy $P(\omega) \geq 0$ and $\sum_{\omega \in \Omega} P(\omega) = 1$.

This is a *linear program*!
- Derived first by Hailperin (1965).

Somewhat surprisingly, NP-complete problem.
- The same as usual satisfiability (!?!)!

Note: solution is at extreme points.
Exercise

Build linear program:

- $P(A) \geq \alpha$.  
- $B \rightarrow C$.  
- $P(B) = \beta$.  

Can you give bounds for $P(A \land B \land C)$?
Solution

\[ P(A) \geq \alpha, \ B \rightarrow C, \ P(B) = \beta. \]

Define:

<table>
<thead>
<tr>
<th>( \omega_i )</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Then \( \omega_3 \) and \( \omega_7 \) are impossible; and

\[ p_5 + p_6 + p_8 = \alpha, \quad p_4 + p_8 = \beta, \quad p_i \geq 0, \quad \sum_i p_i = 1. \]
de Finetti’s fundamental theorem

Given $m$ assessments over events $H_i$, is there a probability measure over them?

And how about the allowed assessments over another event $H_0$?

Theorem: $P(H_0)$ belongs to an interval with constraints given by other assessments

(and the usual $P(\omega) \geq 0$ and $\sum_{\omega \in \Omega} P(\omega) = 1$).

This is a linear program.

Well, this is the same linear program as before (Gilio (1980)).
Exercise

Coletti and Scozzafava (1999).

- Take $H_1$, $H_2$, $H_3$.
- Assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program.
Exercise

Coletti and Scozzafava (1999).

- Take $H_1, H_2, H_3$.
- Assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program.

- $x_1 = P(A_1)$; $A_1 = H_1 \cap H_2 \cap H_3^c$.
- $x_2 = P(A_2)$; $A_2 = H_1 \cap H_2^c \cap H_3^c$.
- $x_3 = P(A_3)$; $A_3 = H_1^c \cap H_2 \cap H_3^c$.
- $x_4 = P(A_4)$; $A_4 = H_1^c \cap H_2 \cap H_3$.
- $x_5 = P(A_5)$; $A_5 = H_1^c \cap H_2^c \cap H_3^c$. 
Exercise

Coletti and Scozzafava (1999).

- Take $H_1, H_2, H_3$.
- Assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program.

\[
\begin{align*}
    x_1 + x_2 &= 1/2 \\
    x_1 + x_3 + x_4 &= 1/5 \\
    x_4 &= 1/8 \\
    x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\
    x_1 \geq 0, & \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0.
\end{align*}
\]
Conditional probabilities

- Assessment $P(A|B) \geq \alpha$.

- Transform to (Hailperin (1965) and many others later):

$$P(A \land B) \geq \alpha P(B).$$

- Or use the language of events.

- Still a linear program!
Exercise

Coletti and Scozzafava (1999).

- Take $H_1, H_2, H_3$.
- Assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2, P(H_2) = 1/5, P(H_3) = 1/8$.
- Also, $P(H_2|H_1 \cup H_2) \geq 1/2$.

Build linear program.
Column generation

- Probabilistic satisfiability is
  \[
  \min 0p \\
  \text{subject to} \quad Ap \geq \alpha, p \geq 1.
  \]

- General problem minimizes \( cp \).

- The difficulty is that \( p \) has \( 2^n \) elements (for a problem with \( n \) propositions).

- The usual technique is \textit{column generation}.
  - That is, generate only those columns of \( A \) that are necessary
  - (at any given time, simplex only needs \( m \) columns where \( m \) is number of lines of \( A \)).
The mechanics of column generation

Use the revised simplex algorithm.
- That is, keep only a basis \((m \times m)\).
- Must decide whether to bring a column into the basis.

Then choose the column using a nonlinear subproblem:
- Solve \(\min_j c_B A_B^{-1} A_j\).
- Note that \(A_j\) contains a set of logical formulas.
- This is a MAXSAT problem.
- Replace:

\[
X \land Y \equiv XY, \quad X \lor Y \equiv X + Y - XY, \quad \neg X \equiv 1 - X.
\]

- It can be reduced to linear (integer) programming!
Integer programming

Very useful fact:

Consider product $a \times b$, where

- $a \in [0, 1]$.
- $b$ is either 0 or 1.

Create a new variable $c$, replace $a \times b$ by $c$ and add

$$0 \leq c \leq b;$$

$$a - 1 + b \leq c \leq a.$$  

Now solve by linear (integer) programming!
PSAT with column generation

- Best results in the literature: hundreds of propositions, hundreds of assessments (Perron et al 2004), using lots of special tricks.

- There are also a few special cases that are “easy” and several variants, etc.
  - For instance, when formulas can be put in a “tree” structure (Andersen & Pretolani 1999).
  - Also if formulas can be organized in junction trees (van der Gaag 1991).

- (Also, approximation methods based on local search for large problems, but really no guarantees yet...)
Phase transitions?

![Graph showing phase transitions with lines for n=50, k=3, n=50, k=4, and n=50, k=5.](image)
Aside: PPL system

- Interface in Python, connects to CPLEX or free linear programming tools (at http://www.pmr.poli.usp.br/ltd/Software/PPL/index.html).

```python
>>> s1 = 'a <=> (b?c)'
>>> s1
'a <=> (b?c)'
>>> s2 = PPL.toCNF(s1)
>>> s2
'((?b j a) & (?c j a) & (b j c j ?a))'
>>> PPL.p(s1, 0.5)
>>> s3 = 'd j (e & f) j g'
>>> PPL.p(s3, 0.3, 0.8)
>>> PPL.checkCoherence()
Coherent!
```

- Another package by Dickey (see SIPTA Newsletter).
Computing conditional probabilities

Now suppose we wish $P(A|B) = \min P(A|B)$.

This is not a linear program (it is a linear fractional program).

However, it can be solved through linear programming:
- Charnes-Cooper transformation (similar solutions by White, Snow).
- Dinkelbach-Jagannathan algorithm (similar solutions by Walley, Lavine).
Charnes-Cooper transformation

Wish to solve:

\[
\min_p \frac{\sum_i f_i \alpha_i p_i}{\sum_i \alpha_i p_i} \quad \text{s.t.} \quad Ap \geq 0, \sum_i p_i = 1, p_i \geq 0.
\]

where \( \sum_i \alpha_i p_i > 0 \).

Change variables to

\[
q_i = \frac{p_i}{\sum_i \alpha_i p_i}.
\]

Now:

\[
\min_q \sum_i f_i \alpha_i q_i \quad \text{s.t.} \quad Aq \geq 0, \sum_i \alpha_i q_i = 1, q_i \geq 0.
\]
Exercise

- Take $H_1$, $H_2$, $H_3$.
- Assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program to compute $P(H_1|H_1 \cup H_2)$, applying the Charnes-Cooper transformation.
Solution

- Take $H_1$, $H_2$, $H_3$, assume $H_3 \subset H_1^c \cap H_2$.
- Assessments $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program to compute $P(H_1|H_1 \cup H_2)$.

First,

$$\min \frac{x_1 + x_2}{x_1 + x_2 + x_3 + x_4} \quad \text{s.t.}$$

$$x_1 + x_2 = 1/2; \quad x_1 + x_3 + x_4 = 1/5; \quad x_4 = 1/8; \quad x_i \geq 0; \quad \sum x_i = 1.$$

Then

$$\min \frac{x_1 + x_2}{x_1 + x_2 + x_3 + x_4} \quad \text{s.t.}$$

$$x_1/2 + x_2/2 - x_3/2 - x_4/2 - x_5/2 = 0; \quad 4x_1/5 - x_2/5 + 4x_3/5 + 4x_4/5 - x_5/5 = 0;$$
$$-x_1/8 - x_2/8 - x_3/8 + 7x_4/8 - x_5/8 = 0; \quad x_i \geq 0; \quad \sum x_i = 1.$$
Solution

- **Take** $H_1$, $H_2$, $H_3$, assume $H_3 \subset H_1^c \cap H_2$.
- **Assessments** $P(H_1) = 1/2$, $P(H_2) = 1/5$, $P(H_3) = 1/8$.

Build linear program to compute $P(H_1|H_1 \cup H_2)$.

First,

$$\min \left( \frac{x_1 + x_2}{x_1 + x_2 + x_3 + x_4} \right) \quad \text{s.t.}$$

$$x_1 + x_2 = 1/2; \quad x_1 + x_3 + x_4 = 1/5; \quad x_4 = 1/8; \quad x_i \geq 0; \quad \sum x_i = 1.$$ 

Then

$$\min (y_1 + y_2) \quad \text{s.t.}$$

$$y_1/2 + y_2/2 - y_3/2 - y_4/2 - y_5/2 = 0; \quad 4y_1/5 - y_2/5 + 4y_3/5 + 4y_4/5 - y_5/5 = 0;$$

$$-y_1/8 - y_2/8 - y_3/8 + 7y_4/8 - y_5/8 = 0; \quad y_i \geq 0; \quad \sum_{i=1}^{4} y_i = 1.$$
Larger example (based on Jaeger 1994)

Take:
- AntarticBird → Bird,
- FlyingBird → Bird,
- Penguin → Bird,
- FlyingBird → Flies,
- Penguin → ¬Flies,
- \( P(\text{FlyingBird}|\text{Bird}) = 0.95 \),
- \( P(\text{AntarticBird}|\text{Bird}) = 0.01 \),
- \( P(\text{Bird}) \geq 0.2 \),
- \( P(\text{FlyingBird} \lor \text{Penguin}|\text{AntarticBird}) \geq 0.2 \),
- \( P(\text{Flies}|\text{Bird}) \geq 0.8 \).

Then
- \( P(\text{FlyingBird}|\text{Bird} \land \neg \text{AntarticBird}) \in [0.949, 0.960] \),
- \( P(\text{Penguin}|\neg \text{AntarticBird}) \in [0.000, 0.050] \).
Note:

\[
\lambda = \min \frac{P(A \cap B)}{P(B)},
\]

iff

\[
\min (P(A \cap B) - \lambda P(B)) = 0,
\]

assuming \( P(B) > 0 \).

The left side is strictly decreasing function of \( \lambda \).

So, we can bracket \( \lambda \).
Also,

\[ \lambda = \min \frac{E[f(X)B]}{P(B)}, \]

iff

\[ \min (E[f(X)B] - \lambda P(B)) = 0 \]

or, rather,

\[ \min E[(f(X) - \lambda)B] = 0; \]

that is,

\[ E[(f(X) - \lambda)B] = 0. \]

This is Walley’s Generalized Bayes Rule (GBR).

Walley proposed iteration:

\[ \mu_{i+1} = \mu_i + 2E[(f(X) - \mu_i)B] / (\overline{P}(B) + P(B)). \]
Lavine’s algorithm

In 1991, Lavine published a paper on robust statistics with the same algorithm, apparently unaware of the literature.

Lavine’s algorithm became quite popular.

Until Lavine’s algorithm, calculation of posterior lower expectations in robust statistics usually relied on very special arguments.

Often, minimax theory.
Now, imprecise likelihoods

Suppose we have $K(X)$ (“prior”) and $K(Y|X = x)$ for each $x$ (“likelihood”).

Suppose $K(Y|X = x)$ is *separately specified* (important condition!).

If $P(Y = y) > 0$, $E[f(X)|Y = y]$ is the unique solution of the equation

$$E[(f(X) - \lambda)p_\lambda(y|X)] = 0,$$

where

$$p_\lambda(y|X) = \begin{cases} 
E[y|x] & \text{if } f(x) \geq \lambda \\
\frac{E[y|x]}{E[y|x]} & \text{if } f(x) < \lambda
\end{cases}$$
Dealing with imprecise likelihoods

\[ E[f(X)|Y = y] = \min_{p', p''} \left[ \frac{\sum_i (f_i L_y(x_i) p'_i + f_i U_y(x_i) p''_i)}{\sum_j (L_y(x_j) p'_j + U_y(x_j) p''_j)} \right], \]

subject to:

\[ A(p' + p'') \leq 0, \]

\[ \sum_i (p'_i + p''_i) = 1, \quad p'_i \geq 0, p''_i \geq 0. \]
Example (based on White 1986)

- Variable with 4 values \( \{\theta_1, \theta_2, \theta_3, \theta_4\} \),

\[
2.5 p(\theta_1) \geq p(\theta_4) \geq 2p(\theta_1),
\]

\[
10p(\theta_3) \geq p(\theta_2) \geq 9p(\theta_3), \quad p(\theta_2) = 5p(\theta_4).
\]

- Also, bounds on likelihood:

\[
L(x|\theta_1) = 0.9, \quad L(x|\theta_2) = 0.1125,
\]
\[
L(x|\theta_3) = 0.05625, \quad L(x|\theta_4) = 0.1125,
\]
\[
U(x|\theta_1) = 0.95, \quad U(x|\theta_2) = 0.1357,
\]
\[
U(x|\theta_3) = 0.1357, \quad U(x|\theta_4) = 0.1357.
\]
Example: solution

\[ P(\theta_1|x) = \min_{p', p''} (0.9p_1' + 0.95p_1''), \]
\[ p' \geq 0, \quad p'' \geq 0, \]
\[
\begin{bmatrix}
-\frac{5}{2} & 0 & 0 & 1 \\
2 & 0 & 0 & -1 \\
0 & -1 & 0 & 5 \\
0 & 1 & 0 & -5 \\
0 & -1 & 9 & 0 \\
0 & 1 & -10 & 0
\end{bmatrix}
\]
\[ [p' + p''] \leq 0, \]
\[ F_1 \alpha' + F_2 \alpha'' = 1, \text{ where} \]
\[ F_1 = [0.9, 0.1125, 0.0562, 0.1125], \quad F_2 = [0.95, 0.1357, 0.1357, 0.1357]. \]
By linear programming: \( P(\theta_1|x) = 0.2881. \)
Independence relations

1. We may easily face some “inferential vacuity”: $A$ and $B$ have no logical relation, $P(A) = 1/2$, $P(B) = 1/2$; then $P(A \land B) \in [0, 1/2]$.

2. Introduce independence to reduce inferential vacuity...

   - $A$ and $B$ independent, $P(A) = 1/2$, $P(B) = 1/2$; then $P(A \land B) = 1/4$.

3. Independence leads to
   - nonlinear constraints.
   - open problems concerning complexity.

   - This will take us to credal networks and the like; this is for other talks.
Credal sets

- So far, Boolean and categorical variables, with linear programming.
- Some general terminology and understanding helps.
- A credal set is a set of probability measures (distributions).
- A credal set is usually defined by a set of assessments.

Example:
1. $\Omega = \{\omega_1, \omega_2, \omega_3\}$.
2. $P(\omega_i) = p_i$.
3. $p_1 > p_3$, $2p_1 \geq p_2$, $p_1 \leq 2/3$ and $p_3 \in [1/6, 1/3]$.
4. Take points $P = (p_1, p_2, p_3)$. 
1. $\Omega = \{\omega_1, \omega_2, \omega_3\}$.

2. $P(\omega_i) = p_i$.

3. $p_1 > p_3$, $2p_1 \geq p_2$, $p_1 \leq 2/3$ and $p_3 \in [1/6, 1/3]$.

4. Take points $P = (p_1, p_2, p_3)$.
Baricentric coordinates

The coordinates of a distribution are read on the lines bisecting the angles of the triangle.

\[ P_1 = \left(\frac{2}{3}, \frac{1}{12}, \frac{1}{4}\right) \]
\[ P_2 = \left(\frac{5}{18}, \frac{1}{6}, \frac{5}{9}\right) \]
Exercise

Consider a variable $X$ with 3 possible values $x_1$, $x_2$ and $x_3$. Suppose the following assessments are given:

\[
p(x_1) \leq p(x_2) \leq p(x_3); \]
\[
p(x_i) \geq 1/20 \quad \text{for } i \in \{1, 2, 3\}; \]
\[
p(x_3|x_2 \cup x_3) \leq 3/4.\]

Show the credal set determined by these assessments in baricentric coordinates.
The basics of credal sets

- Credal set with distributions for $X$ is denoted $K(X)$.
- Given credal set $K(X)$:
  - $\underline{E}[X] = \inf_{P \in K(X)} E_P[X]$.
  - $\overline{E}[X] = \sup_{P \in K(X)} E_P[X]$.
- For closed convex credal sets, lower and upper expectations are attained at vertices.
- A closed convex credal set is completely characterized by the associated lower expectation.
  - That is, there is only one lower expectation for a given closed convex credal set.
- The set of conditional distributions from a convex credal set is convex.
Exercise

Suppose the following judgements are stated:

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Desirable

Here “desirable” means $E[X] \geq 0$.

Draw the credal set defined by such assessments.

What can be said about the desirability of

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_3$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$X_4$</td>
<td>-2</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Back to algorithms

- There are details on conditional probabilities that must be analyzed.
- Before, a little more on probabilistic logic: moving to first order.
First-order probabilistic logic

- Now we have constants, relations, functions, quantifiers:
  \[ \text{man}(\text{Socrates}) \lor \text{mortal}(\text{Socrates}) \]
  \[ \forall x : \text{man}(x) \rightarrow \text{mortal}(x). \]

- There are few general techniques here: too many variations.

- Nilsson (1986) advocated: \[ P(\phi) \geq \alpha \] where \( \phi \) is a sentence.
  - This can be solved by linear programming... but there are \textit{decidability} questions.
Example (Jaumard et al 2007)

Assessments:

- \( P(\forall x : \exists y : t(x, y) \land s(y)) = 0.9. \)
- \( P(\exists x : \neg r(x)) = 0.6. \)
- \( P(\exists y : \neg s(y)) = 0.6. \)
- \( P(\forall x : \forall y : \neg t(x, y) \land r(x) \land s(y)) = 0.7. \)

Compute \( P(\exists x : \exists y : \neg t(x, y)) \).

- Only 12 possible worlds (elements in the Lindenbaum algebra).
- Possible to apply linear program; extension to column generation method is open problem.
Other proposals

A different proposal is to impose probabilities over the domain:

“Probability that a randomly selected bird flies is no smaller than 0.9.”

There has been great interest in this kind of probabilistic logic for

- probabilistic logic programming;
- the semantic web;
- probabilistic databases.

Most algorithms are for languages that can be translated to Bayesian networks.

Few general algorithms (good starting point is the work of Thomas Lukasiewicz).
Zero probabilities

- This is one of the most embarrassing challenges in the world of credal sets.

- In the standard theory of probabilities, it is easy to ignore null events (events with probability zero).
  - Such events “will never happen”.
Zero probabilities

- This is one of the most embarrassing challenges in the world of credal sets.
- In the standard theory of probabilities, it is easy to ignore null events (events with probability zero).
  - Such events “will never happen”.
- But now there may be events with zero lower probability and nonzero upper probability.
  - For instance, if $P(B) \leq \alpha$, then $P(B)$ may be zero.
- So, we may observe $A$ and we need to say something about $P(A|B)$.
- This issue has drawn steady interest in the community, but it is not easy to understand.
Zeroes in linear fractional programs

- Note: the linear fractional programs we discussed before required $P(B) > 0$.
- If $P(B) = 0$, then programs become unfeasible.
- They compute:
  \[
  \min E_P[f(X) | B]
  \]
  where $P$ belongs to
  \[
  \{P : P(B) > 0\}.
  \]
Full conditional measures

The most elegant solution is to consider full probability measures.

A full probability measure is a function $P(\cdot|\cdot)$ on $\mathcal{E} \times \mathcal{E}\backslash\emptyset$ where $\mathcal{E}$ is an algebra of events, such that

- $P(A|C) = 1$;
- $P(A|C) \geq 0$ for all $A$;
- $P(A \cup B|C) = P(A|C) + P(B|C)$ when $A \cap B = \emptyset$;
- $P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C)$ when $B \cap C \neq \emptyset$.

Full probability measures allow $P(A|C)$ to be defined even if $P(C) = 0$!
The Krauss-Dubins representation

- We can partition $\Omega$ into events $L_0, \ldots, L_K$, $K \leq N$
- such that the full conditional measure is represented as a sequence of strictly positive probability measures $P_0, \ldots, P_K$, where the support of $P_i$ is restricted to $L_i$.
- $P(A|B) = P(A|B \cap L_B)$, where $L_B$ is the “layer” where $B$ has nonzero probability.
- This representation has been advocated by Coletti & Scozzafava.

Example (note: $P(A) = 0$, but $P(B|A) = \beta$):

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$A^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>$[\beta]_1$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$B^c$</td>
<td>$[1 - \beta]_1$</td>
<td>$1 - \alpha$</td>
</tr>
</tbody>
</table>
Exercise

Consider assessments:

- $P(A) \geq 1/2$.
- $P(A^c \cap B^c) = 1/2$.
- $P(C|A^c \cap B) = 1/3$.

What is the Krauss-Dubins representation?

What is $P(C|B)$?

What is $P(C^c|A^c \cap B)$?
Coletti-Scozzafava’s method

- Run the usual linear program with assessments $P(A_i | B_i) \geq \alpha_i$.

- If all $B_i$ have $P(B_i) > 0$ for all feasible solutions, stop (solution has been found).

Otherwise:

- Collect those $B_i$ with $P(B_i) = 0$ for all feasible solutions.

- Then build another linear program only with those assessments with these $B_i$.

- Repeat until there are no more assessments (inference is vacuous).
Improving the algorithm

Coletti-Scozzafava’s method has been optimized and expanded by Vantaggi, Capotorti and others.

Idea is to quickly detect/exploit zero probabilities.

Check coherence (CkC) package:
http://www.dipmat.unipg.it/~upkd/paid/software.html

Vantaggi has dealt with independence as well.

Overall, many tests to make, to detect whether events may are null.
Other approaches

- Sequence of $2m$ direct linear programs in the worst case (Walley, Pelessoni, Vicig (1999, 2004)).
  - But still, necessary to run additional linear programs to check whether to proceed.
  - Possible to divide number of linear programs by 2, by examining slack variables (Cozman 2002).

- All of this is to check “coherence” in a strong sense.
  - There are weaker concepts of “coherence”.

Algorithms for Imprecise Probability Part I – p. 56/7
Changing gears: Classes of credal sets

- General assessments are flexible (too flexible?) but are hard to handle for general spaces.
- Possible strategy is to focus on a few canonical ways to define credal sets.
- There are many!
  - Neighborhoods.
  - Capacities.
  - Boxes.
- A great deal of this work is found in the literature on robust statistics.
  - Usually, some mix of linear fractional programming (Dinkelbach-Jagannathan algorithm), minimax theory, and creativity with particular problems.
The classic $\epsilon$-contaminated

- Credal set based on $P_0$ and $\epsilon \in (0, 1)$:

$$\{(1 - \epsilon)P_0 + \epsilon Q : \text{any } Q\}.$$ 

- Old model, originally from robust frequentist statistics (Tukey, then Huber).
Exercise

- If $K(X)$ is an $\epsilon$-contaminated class, what are $E[f(X)]$, $\overline{E}[f(X)]$?

- If $P_0$ is always nonzero, what is $\underline{P}(A|B)$, $\overline{P}(A|B)$?

- If one gives a measure $L$ such that $L(\Omega) < 1$, is this an $\epsilon$-contaminated class? If so, what are $P_0$ and $\epsilon$?
Solution

- If $K(X)$ is an $\epsilon$-contaminated class,

$$E[f(X)] = (1-\epsilon)E_0[f(X)], \quad \overline{E}[f(X)] = (1-\epsilon)E_0[f(X)] + \epsilon.$$ 

- If one gives a measure $L$ such that

$$L(\Omega) < 1,$$

this an $\epsilon$-contaminated class

$$\{(1-\epsilon)(L/L(\Omega)) + \epsilon Q\},$$

where $\epsilon = 1 - L(\Omega).$
Other neighborhoods

- Total variation class:
  \[ \{ P : |P(A) - R(A)| \leq \epsilon \}. \]

  (Exercise: Find lower/upper probabilities for event \( A \).)

- Neighborhoods for several metrics; with several contaminations (given moments, given quantiles, given modes); from conjugate families (well-known example is Imprecise Dirichlet Model).

- Bose (1994): several contaminations at once,
  \[ \{(1 - \epsilon)P + \epsilon_1 q_1 + \cdots + \epsilon_n q_n : q_i \in K_i \}. \]
Density bounded classes

- Given two measures $L$ and $U$ such that

$$L \leq U, \quad L(\Omega) \leq 1, \quad U(\Omega) \geq 1,$$

take the set

$$\{P : L \leq P \leq U\}.$$  

- Lower/upper probabilities are easy to compute. For instance,

$$P(A) = \max(L(A), 1 - U(A^c)).$$

- *Constant* bounded class if $kL = P_0 = U/k$ for some $P_0$, $k > 1$. 
Density ratio classes

Given two measures \( L \) and \( U \) such that \( L(A) \leq U(A) \) for every event \( A \),

\[
\{ P = \mu/\mu(\Omega) : L \leq \mu \leq U \}.
\]

That is, you “draw” \( \mu \) between \( L \) and \( U \), then normalize it.

Equivalent definition: set of distributions such that for every \( A \) and \( B \),

\[
\frac{L(A)}{U(B)} \leq \frac{P(A)}{P(B)} \leq \frac{U(A)}{L(B)}.
\]
Facts about density ratio classes

Posterior probability:

\[
P(A|B) = \frac{L(A \cap B)}{L(A \cap B) + U(A^c \cap B)},
\]

\[
\overline{P}(A|B) = \frac{U(A \cap B)}{U(A \cap B) + L(A^c \cap B)}.
\]

Posterior from single likelihood: just multiply \( L \) and \( U \) by likelihood.

There are bracketing algorithms for computing lower/upper expectations.
Constant density ratio class

- Set of distributions $P$ such that

$$\frac{P(A)}{P(B)} \leq \alpha \frac{P_0(A)}{P_0(B)},$$

for distribution $P_0$ and $\alpha > 1$.

- Class is preserved by conditioning/marginalization!
One great (obscure) idea

Wasserman and Kadane (1982) observed that for some classes (total variation, constant bounded, constant ratio), it is possible to sample from the “center” $P_0$ of the neighborhood, and compute lower expectations.

One of the few cases where a sampling algorithm has been applied to credal sets.

It would be nice to see other sampling methods, but hard to imagine how to do it.
And 2-monotone capacities

- If a credal set satisfies
  \[ P(A \cup B) \geq P(A) + P(B) - P(A \cap B), \]
  it is 2-monotone.

- Examples: \( \epsilon \)-contaminated, total variation, density bounded.

- Define \( \overline{F}_X(x) = \overline{P}(X \leq x) \); then
  \[ E[X] = \int_{-\infty}^{\infty} x \, d\overline{F}_X(x). \]

- Also,
  \[ P(A|B) = \frac{P(A \cap B)}{P(A \cap B) + \overline{P}(A^c \cap B)}. \]
And belief functions

- A capacity that is infinitely monotone; that is, for any \( n \),

\[
P(\bigcup_{i=1}^{n} A_i) \geq \sum_{J \subseteq 1, \ldots, n} (-1)^{|J|+1} P(\bigcap_{i \in J} A_i).
\]

- These are basic entities in Dempster-Shafer theory.

- They can always be expressed as a probability mass assignment and a multi-valued mapping.

- Useful:

\[
E[X] = \sum_{A} m(A) \inf_{\omega \in A} X(\omega).
\]
Probability boxes (p-boxes)

- Take two nondecreasing functions $F$ and $\bar{F}$ such that $F \leq \bar{F}$.

- The set of distributions such that

$$\{ P : F \leq F \leq \bar{F} \}.$$ 

is a p-box.

- There has been work on risk assessment and reliability analysis with p-boxes: often discretization of continuous possibility spaces and then linear programming.
Changing gears: Decision making

- Set of acts $\mathcal{A}$, need to choose one.
  - There are several criteria!

- $\Gamma$-minimax:
  \[
  \arg \max_{X \in \mathcal{A}} \mathbb{E}[X].
  \]

- Maximal: maximal elements of the partial order $\succ$. That is, $X$ is maximal if
  there is no $Y \in \mathcal{A}$ such that $\mathbb{E}_P[Y - X] > 0$ for all $P \in K$.

- $E$-admissibility: maximality for at least a distribution. That is, $X$ is $E$-admissible if
  there is $P \in K$ such that $\mathbb{E}_P[X - Y] \geq 0$ for all $Y \in \mathcal{A}$.

- Maximax, interval dominance, etc.
Comparing criteria

Three acts: \( a_1 = 0.4; a_2 = 0/1 \) if \( A/A^c; a_3 = 1/0 \) if \( A/A^c \).

\[
E[a_i]
\]

\[
P(A) \in [0.3, 0.7].
\]

\( \Gamma \)-minimax: \( a_1 \); Maximal: all of them; E-admissible: \( \{a_2, a_3\} \).
Exercise

Credal set \( \{P_1, P_2\} \):

\[
P_1(s_1) = \frac{1}{8}, \quad P_1(s_2) = \frac{3}{4}, \quad P_1(s_3) = \frac{1}{8},
\]

\[
P_2(s_1) = \frac{3}{4}, \quad P_2(s_2) = \frac{1}{8}, \quad P_2(s_3) = \frac{1}{8},
\]

Acts \( \{a_1, a_2, a_3\} \):

<table>
<thead>
<tr>
<th></th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>(a_2)</td>
<td>2.5</td>
<td>3.5</td>
<td>5</td>
</tr>
<tr>
<td>(a_3)</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Which one to select?
Solution

\[ P_1(s_1) = 1/8, \quad P_1(s_2) = 3/4; \quad P_1(s_3) = 1/8, \quad P_2(s_1) = 3/4, \quad P_2(s_2) = 1/8, \quad P_2(s_3) = 1/8. \]

Acts \( \{a_1, a_2, a_3\} \):

<table>
<thead>
<tr>
<th></th>
<th>( s_1 )</th>
<th>( s_2 )</th>
<th>( s_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>2.5</td>
<td>3.5</td>
<td>5</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Then:

\[ E_1[a_1] = 3/8 + 18/8 + 4/8 = 25/8; \]
\[ E_1[a_2] = 2.5/8 + 21/8 + 5/8 = 28.5/8; \]
\[ E_1[a_3] = 1/8 + 15/8 + 4/8 = 35/8. \]
\[ E_2[a_1] = 18/8 + 3/8 + 4/8 = 25/8; \]
\[ E_2[a_2] = 15/8 + 3.5/8 + 5/8 = 23.5/8 \]
\[ E_2[a_3] = 2/8 + 5/8 + 4/8 = 11/8. \]
A quick discussion

- Limited to finite set of acts.

- Consider $\Gamma$-minimax:
  - Compute $E[a_i]$ for each act.
  - Select act with highest $E[a_i]$.

- (Considerable minimax theory in Berger’s book (1985).)
Maximality

- Find $\Gamma$-minimax solution $a_0$.
- For each other act $a_i \neq a_0$, verify whether
  \[ E_P[a_0 - a_i] \geq 0; \]
  for all $P$; if so, discard $a_i$.
- That is, verify whether
  \[ E[a_0 - a_i] \geq 0. \]
E-admissibility

- For each act $a_i$:
  - Collect all constraints that must be satisfied by $P$.
  - Add constraints
    \[ E_P[a_i - a_j] \geq 0 \]
    for every $a_j \neq a_i$.
  - If all these constraints can be satisfied for some $P$, then $a_i$ is E-admissible.

- This scheme can be extended to problems with mixed acts (Utkin and Augustin 2005).
Credal set $\{P_1, P_2\}$:

\[
P_1(s_1) = \frac{1}{8}, \quad P_1(s_2) = \frac{3}{4}, \quad P_1(s_3) = \frac{1}{8},
\]

\[
P_2(s_1) = \frac{3}{4}, \quad P_2(s_2) = \frac{1}{8}, \quad P_2(s_3) = \frac{1}{8},
\]

Acts $\{a_1, a_2, a_3\}$:

<table>
<thead>
<tr>
<th></th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$a_2$</td>
<td>2.5</td>
<td>3.5</td>
<td>5</td>
</tr>
<tr>
<td>$a_3$</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Which one to select?
And if we take convex hull of credal set?
Solution

\[ P_1(s_1) = 1/8, \quad P_1(s_2) = 3/4, \quad P_1(s_3) = 1/8, \quad P_2(s_1) = 3/4, \quad P_2(s_2) = 1/8, \quad P_2(s_3) = 1/8. \]

Acts \( \{a_1, a_2, a_3\} \):

<table>
<thead>
<tr>
<th></th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>(a_2)</td>
<td>2.5</td>
<td>3.5</td>
<td>5</td>
</tr>
<tr>
<td>(a_3)</td>
<td>1</td>
<td>5</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Consider \( P = \alpha P_1 + (1 - \alpha) P_2 \).

Then:

\[ E_P[a_2 - a_1] = 10\alpha - 3 \geq 0; \quad \alpha \geq 3/10. \]

And:

\[ E_P[a_2 - a_3] = -30\alpha + 17 \geq 0; \quad \alpha \leq 17/30. \]
Goal of this talk: overview of some central ideas without independence.

Basic tool is linear programming (column generation, etc).

- Full conditional measures require special tools.

There are many special kinds of credal sets with associated algorithms: neighborhoods, capacities, etc.

Decision making (several criteria) requires such calculations.