Introduction to Imprecise Probability and Imprecise Statistical Methods

Frank Coolen

UTOPIAE Training School, Strathclyde University

22 November 2017
Classical, precise probability: $P(A)$ for each event $A$ of interest, satisfying Kolmogorov’s axioms

Very successful!

Applied probability, stochastic processes, operations research, statistics, decision support, reliability, risk, et cetera

_Problems?_

Requires a very high level of precision and consistency of information

May be too restrictive to cope carefully with the multi-dimensional nature of uncertainty

Quality of underlying knowledge and information cannot be adequately represented
Cluster analysis comprises a range of methods for classifying multivariate data into subgroups. By organizing multivariate data into such subgroups, clustering can help reveal the characteristics of any structure or patterns present. These techniques have proven useful in a wide range of areas such as medicine, psychology, market research and bioinformatics.

This 5th edition of the highly successful Cluster Analysis includes coverage of the latest developments in the field and a new chapter dealing with finite mixture models for structured data. Real life examples are used throughout to demonstrate the application of the theory, and figures are used extensively to illustrate graphical techniques. The book is comprehensive yet relatively non-mathematical, focusing on the practical aspects of cluster analysis.

Key Features:

- Presents a comprehensive guide to clustering techniques, with focus on the practical aspects of cluster analysis.
- Provides a thorough revision of the fourth edition, including new developments in clustering longitudinal data and examples from bioinformatics and gene studies.
- Updates the chapter on mixture models to include recent developments and presents a new chapter on mixture modelling for structured data.

Practitioners and researchers working in cluster analysis and data analysis will benefit from this book.
Imprecise Probability

Lower Probability $P(A)$ and Upper Probability $\bar{P}(A)$, with $0 \leq P(A) \leq \bar{P}(A) \leq 1$

If $P(A) = \bar{P}(A) = P(A)$ for all events $A$: precise probability

If $P(A) = 0$ and $\bar{P}(A) = 1$: complete lack of knowledge about $A$

For disjoint events $A$ and $B$:

$P(A \cup B) \geq P(A) + \bar{P}(B)$ and $\bar{P}(A \cup B) \leq P(A) + \bar{P}(B)$

$P(\text{not-}A) = 1 - \bar{P}(A)$
Closed convex set $\mathcal{P}$ of precise probability distributions:

$\underline{P}(A) = \inf_{p \in \mathcal{P}} p(A)$ and $\overline{P}(A) = \sup_{p \in \mathcal{P}} p(A)$

**Subjective interpretation:**

$\underline{P}(A)$: *maximum price* at which buying gamble paying 1 if $A$ occurs and 0 else is *desirable*

$\overline{P}(A)$: *minimum price* at which selling this gamble is *desirable*

$\underline{P}(A)$ can be interpreted as reflecting the evidence in favour of event $A$, $1 - \overline{P}(A)$ as reflecting the evidence against $A$, hence in favour of not-$A$

Imprecision $\Delta(A) = \overline{P}(A) - \underline{P}(A)$ reflects lack of perfect information about (the ‘probability’ of) $A$
‘Imprecise Probability’: an unfortunate misnomer as lower and upper probabilities enable more accurate quantification of uncertainty than precise probability

Brings together a variety of different theories (including Dempster-Shafer Theory and Interval Probability)

Successful applications start to appear, e.g. in classification, reliability and risk analysis (e.g. $\underline{P}(A) = 0$ and $\overline{P}(A) > 0$ if $A$ never yet observed)

www.sipta.org
Using Imprecise Probabilities

\( \mathcal{P} \) defined directly (e.g. set of (prior) distributions) or indirectly (e.g. constraints from expert elicitation)

Inference using Imprecise Probabilities: mostly ‘Generalized Bayesian’

Can reflect prior-data conflict

Bounds on inferences as functions of all \( p \in \mathcal{P} \)

Robustness; indecision? May require further criteria to reach final decision

Exploring the wider opportunities still at early stage

Nonparametric Predictive Inference: (‘objective’) frequentist theory

www.npi-statistics.com
Very many research challenges! Both on theory and applications, ideally combined

Includes computational challenges (optimisation!), and e.g. also on simulation

The main challenge: 

*use imprecision to do things better or easier, preferably both!*
Imprecise statistical methods

Example of ‘generalized Bayes’ (standard precise model with set of prior distributions) in next lecture.

I also strongly belief that ‘imprecision’, as a wider idea than ‘imprecise probability’, can be useful for new methods for dealing with uncertainty and information. For example, they may provide new perspectives on ‘machine learning’ methods or even be quite ‘ad hoc’, but possibly relatively simple and useful. Two examples follow..
A robust weighted support vector regression-based software reliability growth model

$n$ training data (examples) $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, in which $x_i \in \mathbb{R}^m$ represents a feature vector involving $m$ features and $y_i \in \mathbb{R}$ is an output variable.

$f(x) = \langle a, \phi(x) \rangle + b$, with parameters $a = (a_1, \ldots, a_m)$, $b$; $\phi(x)$ is a feature map $\mathbb{R}^m \rightarrow G$ such that the data points are mapped into an alternative higher-dimensional feature space $G$; $\langle \cdot, \cdot \rangle$ denotes the dot product.
$$R_{SVR} = \min_{a,b} \max_{w \in P} \left( C \sum_{i=1}^{n} w_i l(y_i, f(x_i)) + \frac{1}{2} \langle a, a \rangle \right)$$

‘Imprecision’ comes into this method through the use of a set of weights $w_i$:

$$w_i \geq (1 - \varepsilon)n^{-1}$$

and

$$w_1 \leq w_2 \leq \ldots \leq w_n$$

and

$$w_1 + \ldots + w_n = 1$$

We use $\varepsilon$ as a further parameter that we choose for ‘optimal performance’ for given data. This method works almost always better than standard approach with defined weights $1/n$. 
Accelerated Life Testing

Constant stress testing

Challenge: transform information from increased stress level such that it is representative for information about failure times under the normal conditions.

The power-Weibull model:

Weibull($\alpha_i$, $\beta$) for failure times at stress level $i = 0$ (normal), 1, 2, . . . , $k$,

$$P(T > t) = \exp \left\{ - \left( \frac{t}{\alpha_i} \right)^{\beta} \right\}$$
Stress level $i$ quantified by positive measurement $V_i$, increasing in $i$

$$
\alpha_i = \alpha \left( \frac{V_0}{V_i} \right)^p
$$

An observation $t^i$ at stress level $i$, so subject to stress $V_i$, can be interpreted as an observation

$$
t^i \left( \frac{V_i}{V_0} \right)^p
$$

at the normal stress level.
Our Approach

1. We estimate (e.g. MLE) the parameters $\alpha, \beta, p$

2. We use the estimate $\hat{p}$ (only!) to transform data to the normal stress level

3. We use all combined (transformed) data at normal stress level for nonparametric predictive inference, providing lower and upper predictive survival functions for the next unit at normal stress level

4. To get further robustness wrt model assumptions and estimation bias, we replace single value of $p$ by an interval; transformed observation becomes an interval

*Ad hoc?*
**Table**: Failure times at three voltage levels.

<table>
<thead>
<tr>
<th>Voltage ( V )</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_0 = 52.5 )</td>
<td>245, 246, 350, 550, 600, 740, 745, 1010, 1190, 1225, 1390, 1458, 1480, 1690, 1805, 2450, 3000, 4690, 6095, 6200*</td>
</tr>
<tr>
<td>( V_1 = 55.0 )</td>
<td>114, 132, 144, 162, 222, 258, 300, 312, 396, 444, 498, 520, 745, 772, 1240, 1266, 1464, 1740*, 2440*, 2600*</td>
</tr>
<tr>
<td>( V_2 = 57.5 )</td>
<td>168, 174, 234, 252, 288, 288, 294, 348, 390, 408, 444, 510, 528, 546, 558, 690, 696, 714, 900*, 1000*</td>
</tr>
</tbody>
</table>
Figure: Lower and upper survival functions with \( \hat{p} = 15.104 \)
**Figure:** Lower and upper survival functions with $[p, \bar{p}] = [14.5, 15.5]$
**Figure**: Lower and upper survival functions with $[\rho, \bar{\rho}] = [13.0, 17.0]$
Figure: Lower and upper survival functions with \([p, \overline{p}] = [10.0, 20.0]\)
Collaborators

Tahani Coolen-Maturi (Durham)
Lev Utkin (St Petersburg)
Yi-Chao Yin (China, visiting PhD student Durham 2015-16)