Mapping conditional scenarios for knowledge structuring in tail dependence elicitation: A case study on innovation and risk in the UK higher education sector

Christoph Werner | University of Strathclyde, UK | COST Conference, Espoo | 25th -27th April 2017
Need for formal elicitation processes

Why should we use formal elicitation processes?

- The need for a structured and formal elicitation process has been recognised since the earliest approaches of Probabilistic Risk Analysis (PRA):
  - Expert judgement was used only in a semi-formal way in one of the first full-scale PRAs, the original Reactor Safety Study by the US Nuclear Regulatory Commission (USNRC, 1975)
  - A more scientific and transparent elicitation process were developed for the subsequent studies, known as NUREG 1150 (USNRC, 1991; 1987; Keeney and Von Winterfeldt, 1991)

- Another formal elicitation process was proposed by the Stanford Research Institute (SRI).
Structured process for eliciting dependence

### Overview of Dependence Elicitation Process

**Problem Identification:**
- Identify relationships between variables, specify dependence problem/determine modelling context
- Design elicitation for chosen dependence model

**Choice of Elicited Parameters:**
- Account for desiderata of elicited forms
- Consider prevalence of cognitive fallacies for certain forms
- Account for experts' familiarity with dependence parameter

**Preparation of Background Information, Briefing Document and Elicitation Document**

**Expert Identification and Selection**

**Specification of Marginal Distributions:**
- Assess from historical data (if available) or decide whether to assess in same or separate EJ session

**Trial-Run of Elicitation**

**Training and Motivation:**
- Familiarise the expert with elicited form
- Complement feedback of training questions with simulation-based learning approaches
- Explain common biases

**Knowledge and Belief Structuring:**
- Assess experts' rationale behind assessment

**Quantitative Elicitation**

**Aggregation of Expert Judgements:**
- Decide on reasonable aggregation method
- Base probabilistic independence on structural information

**Feedback and Robustness Analysis:**
- Use graphical outputs for “feeding back”

**Documentation**
Main cognitive fallacies for dependence elicitation

### Overview of Biases for Eliciting Dependence

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference(s)</th>
<th>Description</th>
<th>Originates with</th>
<th>Suggested Remedies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confusion of the inverse</td>
<td>Meehl and Rosen (1955), Eddy (1982), Dawes (1988), Hastie and Dawes (2001)</td>
<td>Experts confuse conditional probabilities of ( P(X</td>
<td>Y) ) with its inverse ( P(Y</td>
<td>X) )</td>
</tr>
<tr>
<td>Causality heuristic</td>
<td>Ajzen (1977), Tversky and Kahneman (1980)</td>
<td>Experts overestimate ( P(X</td>
<td>Y) ) when perceiving causal relationship, i.e. ( Y ) causing ( X )</td>
<td>causal interpretation, base-rate neglect</td>
</tr>
<tr>
<td>Insufficiently regressive prediction</td>
<td>Kahneman and Tversky (1973)</td>
<td>Experts translate one scale to the other, not adjusting for imperfect association</td>
<td>representativeness heuristic, predictive interpretation</td>
<td></td>
</tr>
<tr>
<td>Bayesian likelihood bias</td>
<td>Edwards (1965), DaCharme (1970)</td>
<td>Experts are more conservative than Bayes' Theorem implies</td>
<td>representativeness heuristic, base-rate neglect</td>
<td></td>
</tr>
<tr>
<td>Confusion of joint and conditional probabilities</td>
<td>Einhorn and Hogarth (1986)</td>
<td>Experts confuse joint and conditional probabilities</td>
<td>causal/temporal interpretation</td>
<td>address semantic misunderstandings in training, structure rationale/scenarios/functional relationships</td>
</tr>
<tr>
<td>Cell A strategy</td>
<td>Smedslund (1963), Allen (1980), Kao and Wasserman (1993)</td>
<td>Experts overvalue joint presence of variables (in bivariate assessment)</td>
<td>predictive interpretation</td>
<td>clarify underlying assumptions, such as rarity assumption</td>
</tr>
<tr>
<td>Illusory correlation</td>
<td>Chapman and Chapman (1969), Eder et al. (2011)</td>
<td>Experts base assessment on false (pre-existing) belief about relationships</td>
<td>availability bias, causal interpretation</td>
<td>as for availability: provide probability training, counter-examples, relevant statistics (if available) (Montibeller and Von Winterfeldt 2015)</td>
</tr>
</tbody>
</table>

*only main and/or original references listed*
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Brief Overview of UK HE sector

- While general management of the higher education (HE) sector in the United Kingdom has been studied and is well-understood, an aspect that has been neglected in this regard is the management and assessment of risk.

- Considering risks in HE is of key importance as today the sector faces a more complex environment to operate in, for instance:
  - global movement of staff and students together with increased exposure to and reliance on overseas markets
  - variable tuition fees, which increase competition and change students' expectations
  - large investments in infrastructures to facilitate institutional expansion
  - loss of market share due to new technologies
  - And some more general uncertainties affection HE are e.g.: political realities together with national security concerns, such as changing visa requirements, government policies influencing the cost of studies, the alignment and accreditation of degrees and the future impact of e-learning offerings

- A consequence is that higher education institutions need to be able to offer new courses that are in line with the demands of employers and hence manage a variety of courses profitably and effectively within a university department's portfolio.
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Motivation and contribution of case-study

- For decision makers in HE, probabilistic modelling of uncertainties can offer insights into the risks involved with their actions. A common challenge is however that relevant historical data are lacking, in particular for dependence relationships.

- Accounting for and modelling dependence is nevertheless of key importance in such a complex business environment such as HE.

- In this case study, we use expert judgement for quantifying dependence between tuition fee income in 2020/21 for the MSc Business Analysis and Consulting and the MSc Data Science.

- First, we elicit the marginal distribution of both courses’ incomes and then the dependence between them.

- A scenario mapping method is introduced and used with the aim of, supporting experts in making sense of their underlying knowledge and beliefs, and mitigating some common heuristics and biases that might occur when assessing dependence.
Elicitation of Marginal Distributions (1/2)

**T01:** For the academic year 2020/2021, i.e. in 4 years from now, what will be the generated income from the MSc Business Analysis and Consulting?
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Elicitation of Marginal Distributions (2/2)

**T02:** For the academic year 2020/2021, i.e. in 4 years from now, what is the generated income form the MSc Data Science (reminder: this course start from next year onwards)?

![Graph showing income distribution for MSc Data Science 2020/21](image)
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Elicitation of Dependence

After having specified the marginal distributions for our variables of interest, the next step is to assess their dependence. The elicitation questions for this are in the form of conditional probabilities:

1) Given that the generated income of the MSc BAC is below its \textit{median}, i.e. £929,000 in the academic year 2020/21 what is the probability that the MSc Data Science is also below its \textit{median}, i.e £923,000? \( P(DS < 50^{th}|BAC < 50^{th}) \)?

2) Given that the generated income of the MSc BAC is below its \textit{5th quantile}, i.e. £417,000 in the academic year 2020/21 what is the probability that the MSc Data Science is also below its \textit{5th quantile}, i.e £402,000? \( P(DS < 5^{th}|BAC < 5^{th}) \)?
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Conditional Scenario Mapping Process

Process Set-Up
- Experts are made aware of different steps/procedures.
- Formal matters such as confidentiality of personal information are clarified.

Introduction of Specified Quantile for Unconditional Distribution
- The facilitator clarifies the time frame in which scenarios can lie.
- Experts are introduced to specified final condition of required scenarios (based on marginal distributions).

Unstructured Part and Event Classification
- Experts brainstorm for plausible reasons of "being above the specified quantile".
- Experts classify reasons into event types that are introduced by facilitator.

Scenario Generation (1/2)
- Experts as present their events and the facilitator maps them into relevant scenarios with backwards logic.

Peer Feedback
- Experts are given the scenarios of their peers grouped according to the number of experts with the same scenarios, but anonymised.

Scenario Generation (2/2)
- Starting from the trigger events, the scenario is generated. Experts map scenarios for the second distribution.

Assessment of Variable of Interest
- Experts assess the variable of interest with the help of their underlying scenarios.

Possible Assessment Adjustment
- Experts can (but not here to adjust their assessment given the peer feedback.)
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Scenario Definition (1/2)

\[ \omega_i = (x_0, x_1, ..., x_k, x_0, x_1, ..., x_k \in \mathcal{A}_u) \]

- one possible outcome from the entire sample space

- "hypothetical sequences of events constructed as causal chains of argumentation for the purpose of focussing attention on alternative future" (Eden and Ackerman, 1999)

- decomposed into: triggering events, intermediate conditions and specified condition
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Scenario Definition (2/2)

- **Trigger Events (Immediate)**: A trigger event is a plausible initiator of a scenario which is contained in the current state of the world and which may or may not be (fully) observable. Sometimes it might be possible to add words like “start”, “outbreak”, “attack”, “eruption”, “shock” etc. if necessary for clarification; e.g. disease outbreak, terrorist attack, volcanic eruption, oil price shock. If we identify observable trigger events, it should be possible to neglect any events that led to the trigger event as we condition on them anyways. However, if we identify trigger events that are only partly observable, then we need to further include past events which led to the trigger event for ensuring a richer set of scenarios.

- **Trigger Events (Evolving)**: Similar to immediate trigger events, an evolving counter-part exists. These events differ as they describe a longer development which can be seen as an initial cause. It should be possible to insert in a sensible manner words like “development”; e.g. development of (long lasting) rain showers.

- **Enabling Conditions**: Complementary to both types of trigger events, the experts should also identify enabling conditions. In a temporal order these follow from the trigger events and they relate to evolving trends in a system. Therefore, they should include terms such as “higher”/“lower” (or similar), e.g. lower economic growth, higher risk of infection, higher migration.
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Expert B: Unconditional distribution, i.e. MSc BAC ≤ 50th quantile
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Expert B: Conditional distribution, MSc DS < 50th | MSc BAC < 50th
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Expert B: Unconditional distribution, i.e. MSc BAC ≤ 5th quantile
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Expert B: Conditional distribution, MSc DS < 5\textsuperscript{th} | MSc BAC < 5\textsuperscript{th}
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Common independent and lower tail dependent parametric copulas ($rc=0.3$ and $0.7$)
Illustrative case-study: Innovation and risk in the UK Higher Education (HE) sector

Fitting aggregated result to parametric copula

<table>
<thead>
<tr>
<th>Expert</th>
<th>50th quantile</th>
<th>5th quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>Expert 3</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Expert 4</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Expert 5</td>
<td>0.85</td>
<td>0.5</td>
</tr>
<tr>
<td>DM Global</td>
<td>0.787</td>
<td>0.729</td>
</tr>
<tr>
<td>DM EW</td>
<td>0.79</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Graph showing exceedance probability values for different copula models.
Thank you for your attention.