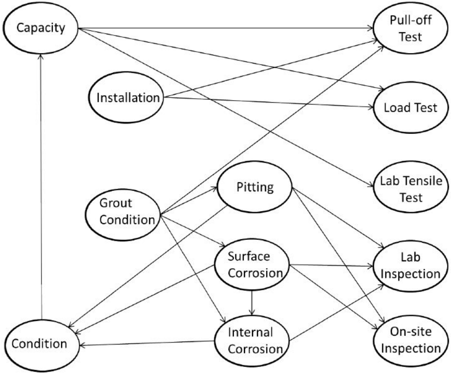
The SHeffield ELicitation Framework and vine copulas in the specification of prior distributions for multinomial models

Kevin Wilson

School of Mathematics and Statistics, Newcastle University, UK

Thanks to Lesley Walls, John Quigley, University of Strathclyde

4th July 2017



- A group of engineers are responsible for a large road bridge.
- The bridge is coming to the end of its useful life.
- The engineers would like to assess the condition of the bridge.

- Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_{m+1})$ and

$$\boldsymbol{Y} \mid (p_1, \ldots, p_{m+1}) \sim \mathrm{Mn}(M, (p_1, \ldots, p_{m+1})).$$

- We wish to define an informative prior distribution over $(p_1, \ldots, p_{m+1})$.
- We would like to include dependency between p_i, p_j for $i \neq j$.
- A complication is the unit sum constraint

$$\sum_{i=1}^{m+1} p_i = 1.$$

News SHELF version 3.0 is a major upgrade (October 2016).

- SHELF offers a formal procedure and resources for conducting elicitation sessions.
- It is primarily focussed on group behavioural elicitation but can be used for individual elicitation.
- The resouces included are:
    - Advice on all aspects of a formal elicitation.
    - Samples of conducted elicitations.
    - Slide sets to use when conducting elicitations.
    - Templates for elicitation records.
    - Software to fit probability distributions to elicited information.
- It can be used to elicit univariate or multivariate prior distributions.

Dirichlet distribution

- Quantity of interest: $\boldsymbol{p} = (p_1, \ldots, p_{m+1})$ where $\sum_{i=1}^{m+1} p_i = 1$.
- Elicited marginal distributions:

$$p_i \sim \mathrm{beta}(a_i, b_i).$$

- Parameter adjustment:

$$a_i^* = \frac{a_i}{\sum_{i=1}^{m+1} \mu_i}, \quad b_i^* = b_i + a_i - a_i^*.$$

- We then need to impose $n_i^* = a_i^* + b_i^* = n$.
- Feedback is given to the experts on the implications of these adjustments.
- No additional elicitations are made beyond those for the marginal distributions.

Gaussian copula elicitation

- The prior distribution takes the form

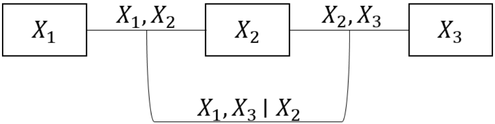
$$f^{(0)}(\boldsymbol{p}) = f_1^{(0)}(p_1) \times \cdots \times f_m^{(0)}(p_m) c(F_1(p_1), \ldots, F_m(p_m)),$$

$$C(u_1, \ldots, u_m) = \Phi_{m,R}\left[\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m)\right],$$

  where $\boldsymbol{p} = (p_1, \ldots, p_m)$, and $u_i = F_i^{(0)}(\theta_i)$.
- The marginal distributions are specified independently from the dependence structure.
- The correlation matrix R is specified by eliciting:

$$P_{i,j} = \Pr(p_i > q_{0.5,i} \text{ AND} p_j > q_{0.5,j}).$$

- Then $r_{i,j} = \sin(\pi P_{i,j} - \pi/2)$.
- The correlation matrix is not guaranteed to be positive semi-definite.



- A bivariate copula is a distribution on $[0,1]^2$, such that,

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) c(F_1(x_1), F_2(x_2)).$$

- A vine is based on the decomposition of a multivariate density into a set of bivariate copulas.
- *Any* vine structure can be used to approximate *any* multivariate distribution to *any* degree of approximation.

$$\begin{aligned} f_{1,2,3}(\boldsymbol{x}) &= f_1(x_1) f_2(x_2) f_3(x_3) \\ &\quad \times c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) \\ &\quad \times c_{13|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)) \end{aligned}$$

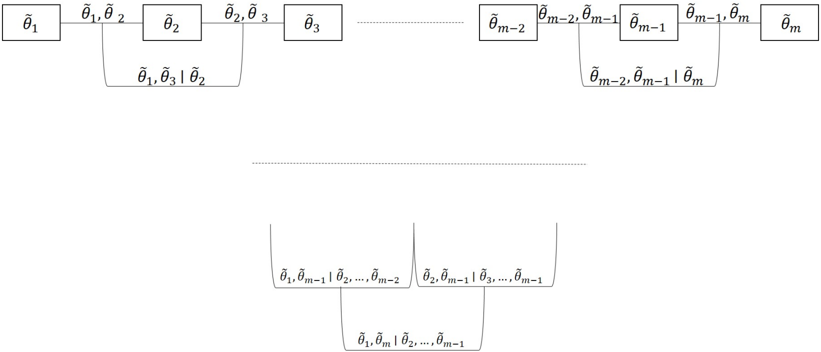
- Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_{m+1})$ and

$$\boldsymbol{Y} \mid (p_1, \ldots, p_{m+1}) \sim \mathrm{Mn}(M, (p_1, \ldots, p_{m+1})).$$

- We would like to specify a flexible multivariate prior over $(p_1, \ldots, p_{m+1})$
- To overcome issues associated with the sum constraint we define:

$$\theta_i = \frac{p_i}{1 - \sum_{j=1}^{i-1} p_j}.$$

- We can easily transform back: $p_i = \theta_i \prod_{j=1}^{i-1}(1 - \theta_i)$, where $(p_1 = \theta_1)$.
- We can then define $f^{(0)}(\boldsymbol{\theta})$ as a D-vine.



- To define the prior distribution for $\boldsymbol{\theta}$, specify:
  1. Marginal prior distributions, $f_i^{(0)}(\theta_i)$.
  2. Unconditional copulas in Tree 1, $c_{i,i+1}$.
  3. Conditional copulas in Trees $2, \dots, m-1$, $c_{i,i+j|i+1,\dots,i+j-1}$.

- First consider the elicitation of marginal distribtions for θ_i.
- Suppose
$$\theta_i \sim \mathrm{beta}(a_i, b_i).$$
- Elicit three quantiles: $(q_{L,i}, m_i, q_{U,i})$.
- Each pair gives exact $(a_{i,j}, b_{i,j})$, $j = 1, 2, 3$ and $(\mu_{i,j}, \sigma^2_{i,j})$.
- We define

$$\begin{aligned} \mu_i &= w_{i,1}\mu_{i,1} + w_{i,2}\mu_{i,2} + w_{i,3}\mu_{i,3}, \\ \sigma_i^2 &= \frac{1}{\sum_{i=1}^{3} w_i^2}(w_{i,1}^2\sigma_{i,1}^2 + w_{i,2}^2\sigma_{i,2}^2 + w_{i,2}^2\sigma_{i,2}^2), \end{aligned}$$

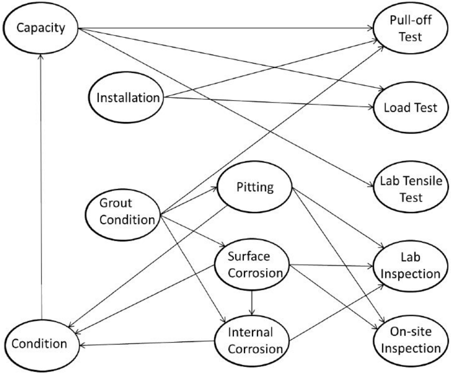
for weights w_i and then

$$a_i = \mu_i^* \left[\frac{\mu_i^*(1-\mu_i^*)}{\sigma_i^2} - 1\right], \; b_i = (1-\mu_i^*)\left[\frac{\mu_i^*(1-\mu_i^*)}{\sigma_i^2} - 1\right].$$

- To elicit dependencies, we condition on some probabilities and ask experts for revised quantiles of other probabilities.
- This would be a challenging task for θ_j and so instead we ask about p_j.
- We can convert quantiles of $p_j \mid p_1, \ldots, p_{j-1}$ into those of $\theta_j \mid \theta_{j-1}$ (Elfadaly and Garthwaite, 2016) via

$$q^*_{k,j} = \frac{q^{'}_{k,j}}{1 - \sum_{l=1}^{j-1} q^{\#}_{0.5,l}},$$

- We condition in each case on $p_i = q^{\#}_{0.5,i}$.
- We fit various bivariate copulas to these three specifications using least squares.
- We can then choose the best fitting copula.
- In subsequent trees in the vine, we condition on the values of more than one probability.



State	$q_{L,i}$	$q_{0.5,i}$	$q_{U,i}$
Original level	0.7	0.75	0.8
Acceptable reduced capacity	0.6	0.7	0.75
Unacceptable reduced capacity	0.2	0.25	0.3
Failed	1	1	1

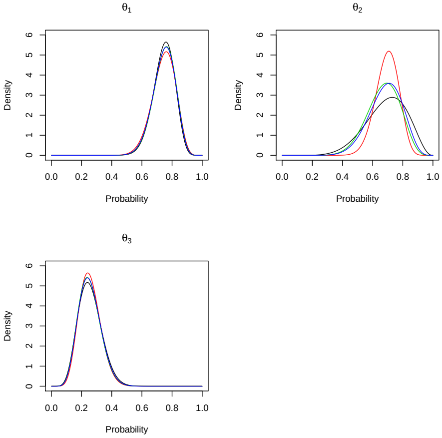


Figure: The marginal prior beta distributions based on each pair of elicited quantiles and the final distribution for $(\theta_1, \theta_2, \theta_3)$. The colours represent $(a_{1,i}, b_{1,i})$ (black), $(a_{2,i}, b_{2,i})$ (red), $(a_{3,i}, b_{3,i})$ (green) and (a_i, b_i) (blue).

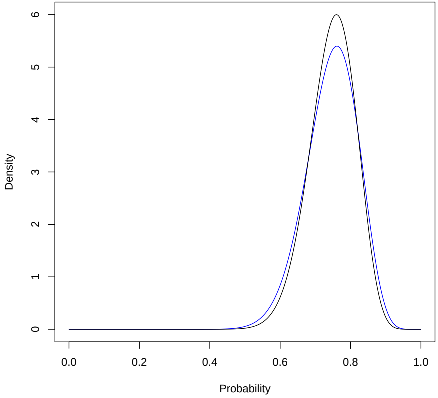


Figure: The marginal prior distribution for θ_1 based on the SHELF (black) and the mean and variance (blue) approach.

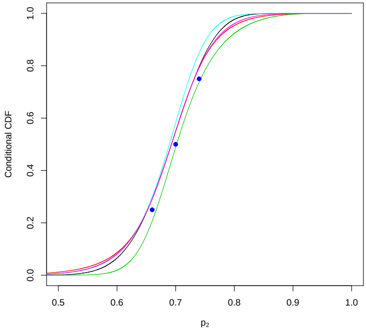


Figure: The conditional CDF of θ_2 given $\theta_1 = q_{0.5,1}$ for the Gaussian copula (black), Frank copula (red), Clayton copula (green), Gumbel copula (light blue) and t-copula (pink), as well as the three elicited quartiles (dark blue)

- Typical approaches to eliciting priors for multinomial distributions restrict the possible dependence structures.
- Vines can give a more flexible dependence specification, with the same number of expert specifications.
- D-vines represent a suitable vine structure and parametric copulas contain the flexibility for the required dependency.
- The elicitation can be expressed in terms of quantities about which we could ask an expert.

"I know nothing about the subject,
but I'm happy to give you my expert opinion."

- An elicitation session using SHELF can involve a number of important roles.
  1. non-specialist: client, co-ordinator
  2. statistical: facilitator, recorder, analyst
  3. domain: experts, advisor
- Stages of an elicitation conducted through SHELF:

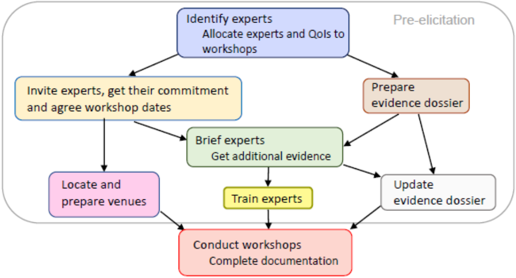


Figure: http://www.tonyohagan.co.uk/shelf/

- The elicitation session itself is broken up into three parts.
  1. Context: purpose, training, expertise, declarations of interest, evidence and definitions.
  2. Univariate distribution: Plausible range, elicitation, fitting (individual and group) plus group discussion and feedback.
  3. Multivariate distribution: Similar to 2. but typically only at group level. Two method available: Dirichlet and Gaussian copula.
- For the univariate elicitations, facilitators choose between three methods: quartile, tertile and roulette.
- The group elicitation concerns the beliefs on an Independent Rational Observer.

- p: the proportion of students achieving grade A* in their GCSE Mathematics at Monkseaton High School in 2016.
- In the quartile and tertile methods we elicit $(q_L, q_{0.5}, q_U)$, three quantiles of p.

Lower quartile Q1

- You should judge it to be equally likely that the true value of X is below Q1 or between Q1 and M
  - Q1 should of course be between your L and M, but it should be closer to M

- In the roulette method, a number of counters called "probs" are placed into different "bins".