Imprecise Inference for 2×2 Tables

Miķelis Bickis
with Naeima Ashleik

University of Saskatchewan

Workshop on Principles and Methods of Statistical Inference with Interval Probabilities
Durham, UK
6 September 2016

University of Saskatchewan

Research supported by NSERC CRSNG

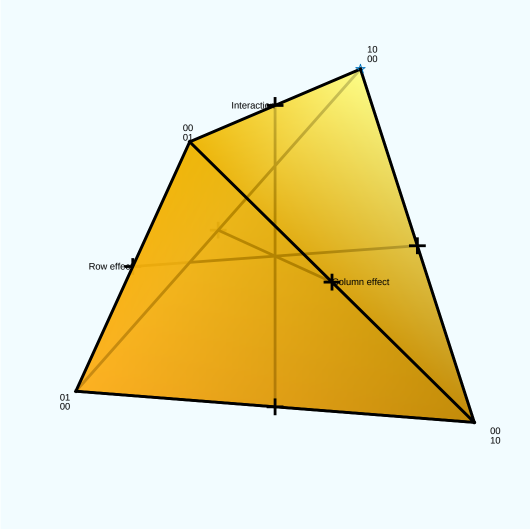
One can parametrize the multinomial distribution for a 2×2 table with cell probabilities

p_{00}	p_{01}
p_{10}	p_{11}

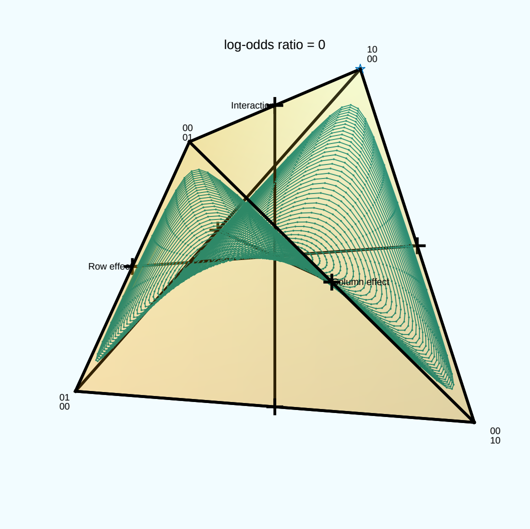
.

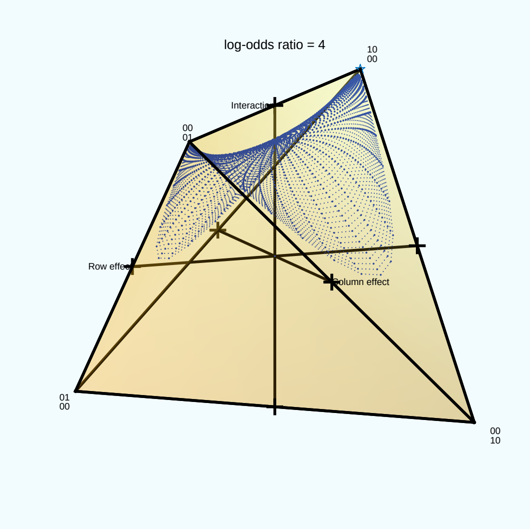
Likelihood arguments can be based on the idea of a single multinomial observation y_{ij} which indicates which of the four cells is observed. The likelihood for n independent observations would be just the product of the likelihoods of the observations, which would be of the same form.

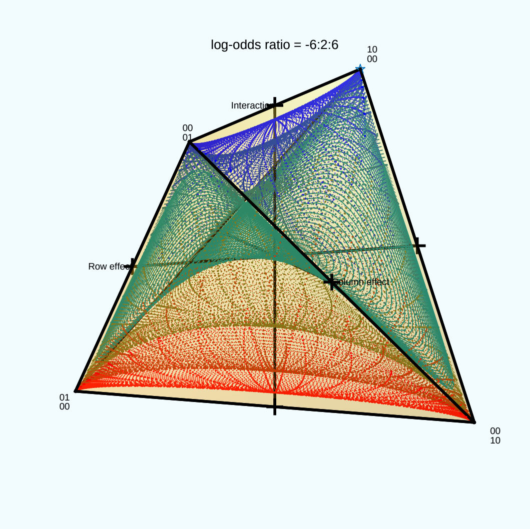
Does it make sense to talk about (classical) independence using imprecise probabilities?



10
00
Interaction
00
01
Row effect
Column effect
01
00
00
10


log-odds ratio = 0
10
00
Interaction
00
01
Row effect
Column effect
01
00
00
10


log-odds ratio = 4
10
00
Interaction
00
01
Row effect
Column effect
01
00
00
10


log-odds ratio = -6:2:6
10
00
Interaction
00
01
Row effect
Column effect
01
00
00
10

What can be said about the geometry of the various kinds of independence/irrelevance properties in the theory of imprecise probability?
Epistemic irrelance, epistemic independence, strong independence etc.?

Let us now reparametrize to:

$$\theta_1 = \log \sqrt{\frac{p_{10}p_{11}}{p_{00}p_{01}}} \tag{1}$$

$$\theta_2 = \log \sqrt{\frac{p_{01}p_{11}}{p_{00}p_{10}}} \tag{2}$$

$$\theta_3 = \log \sqrt{\frac{p_{00}p_{11}}{p_{01}p_{10}}}. \tag{3}$$

Note that $2\theta_3$ is the log odds ratio which is zero in the case of independence.

The inverse transformation then becomes:

$\dfrac{1}{1+e^{\theta_1-\theta_3}+e^{\theta_2-\theta_3}+e^{\theta_1+\theta_2}}$	$\dfrac{e^{\theta_2-\theta_3}}{1+e^{\theta_1-\theta_3}+e^{\theta_2-\theta_3}+e^{\theta_1+\theta_2}}$
$\dfrac{e^{\theta_1-\theta_3}}{1+e^{\theta_1-\theta_3}+e^{\theta_2-\theta_3}+e^{\theta_1+\theta_2}}$	$\dfrac{e^{\theta_1+\theta_2}}{1+e^{\theta_1-\theta_3}+e^{\theta_2-\theta_3}+e^{\theta_1+\theta_2}}$

Denote the observations of the table as:

y_{00}	y_{01}
y_{10}	y_{11}

.

With a single observation, only one of the cell entries would be 1, the others being zeros.

Let's centre the observations with the new variables:

$$\ell_1 = y_{10} + y_{11} - \tfrac{1}{2} \tag{4}$$
$$\ell_2 = y_{01} + y_{11} - \tfrac{1}{2} \tag{5}$$
$$\ell_3 = y_{00} + y_{11} - \tfrac{1}{2}, \tag{6}$$

from which it follows that

$$\begin{array}{|c|c|}\hline y_{00} & y_{01} \\ \hline y_{10} & y_{11} \\ \hline \end{array} = \begin{array}{|c|c|}\hline \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{4} \\ \hline \end{array} + \begin{array}{|r|r|}\hline -\frac{1}{2}\ell_1 - \frac{1}{2}\ell_2 + \frac{1}{2}\ell_3 & -\frac{1}{2}\ell_1 + \frac{1}{2}\ell_2 - \frac{1}{2}\ell_3 \\ \hline \frac{1}{2}\ell_1 - \frac{1}{2}\ell_2 - \frac{1}{2}\ell_3 & \frac{1}{2}\ell_1 + \frac{1}{2}\ell_2 + \frac{1}{2}\ell_3 \\ \hline \end{array}.$$

Thus the ℓ_j variables quantify the deviation of the observation from the uniform expected value of $\frac{1}{4}$ in all cells.
Now, we can write

$$\log p_{ij} = \ell_1\theta_1 + \ell_2\theta_2 + \ell_3\theta_3 - \phi(\theta), \quad i,j = 1,2, \tag{7}$$

where

$$\begin{aligned}\phi(\theta) &= -\tfrac{1}{4}\log\prod_{ij} p_{ij} &(8)\\ &= \log\left(1 + e^{\theta_1-\theta_3} + e^{\theta_2-\theta_3} + e^{\theta_1+\theta_2}\right) - \tfrac{1}{2}(\theta_1+\theta_2-\theta_3). &(9)\end{aligned}$$

Now, from (7) we can see that the distributions of the 2×2 table form an exponential family, with the θ's being canonical parameters and the ℓ's being minimal sufficient statistics. Note that $2\theta_3$ is in fact the log-odds ratio.

Now, if we put a Dirichlet prior on the p_{ij}'s, this will induce a prior on the the θ_j's, and indeed it will be conjugate (in the sense of Diaconis and Ylvisaker).
An imprecise Dirichlet prior will similarly induce an imprecise prior on the θ_j's. We might be particularly interested in upper and lower posterior expectations of the θ_3, which is half the log odds ratio.

If the Dirichlet prior is parametrized in the usual fashion in terms of a concentration parameter s, and marginal expectations t_{ij}, then the posterior expectation of the log odds ratio can be expressed as

$$\psi(y_{00} + st_{00}) - \psi(y_{01} + st_{01}) - \psi(y_{10} + st_{10}) + psi(y_{11} + st_{11})$$

where ψ is the digamma function

$$\psi(x) = \frac{d}{dx} \log \int_0^\infty u^{x-1} e^{-u}\, du.$$

By evaluating this expression for t_{ij} over the simplex, one can find upper and lower posterior expectations.
Will these occur at the extreme points of the simplex?

Suppose now that we put a multivariate normal prior on the θ's? What can we say about the posterior distribution? In particular, what can we say about the posterior marginal distribution of θ_3? Can we put an imprecise prior on the θ's such that we have prior ignorance on θ_3 but allowing learning from data?

What can we say about convexity of sets of posterior distributions?

- A one-dimensional exponential family is stochastically monontone.
- This means that the posterior CDF's corresponding to sets in an interval of hyperparameters will be bracketed by the CDF's at the end points.
- Thus the extreme points of the hyperparameter set will define a P-box.
- This cannot be automatically generalized to multidimensional families because the one-dimensional marginals of an exponential family do not necessarily form an exponential family.

- Under what conditions will extreme points of hyperparmeter sets define extreme points (in the sense of stochastic ordering) of posterior distributions?
- Are there problems of interpretation when this is not the case?
- Is this the case for posterior distributions of log-odds ratio?