Priors and Inferences for Two or More Proportions

Malcolm Farrow
(Joint work with Kevin Wilson)

October 2010

Contents

Contents

Contents

Contents

Contents

1. Introduction.
2. Design problem.
3. Measures of association.
4. Some possible joint belief structures.
    (i) Full probability distributions.

Contents

Simple Motivational Example: 2×2 contingency table.
Two binomial distributions (one fixed margin).

	Outcome 0	Outcome 1
Group 1	$n_1 - Y_1$	Y_1
Group 2	$n_2 - Y_2$	Y_2

Given θ_1, θ_2

$$
\begin{aligned}
Y_1 &\sim Bin(n_1,\ \theta_1) \\
Y_2 &\sim Bin(n_2,\ \theta_2)
\end{aligned}
$$

Introduction

2×2 contingency table

- Very simple.
- Only two parameters.
- No real difficulty with numerical calculations when we use non-conjugate priors.

2 × 2 contingency table

- Very simple.
- Only two parameters.
- No real difficulty with numerical calculations when we use non-conjugate priors.
- E.g.

$$\eta_i = g(\theta_i)$$

$g()$: logit, probit, whatever
(η_1, η_2): bivariate normal prior

BUT . . .

BUT ...

1. Generalisation to bigger problems.

BUT ...

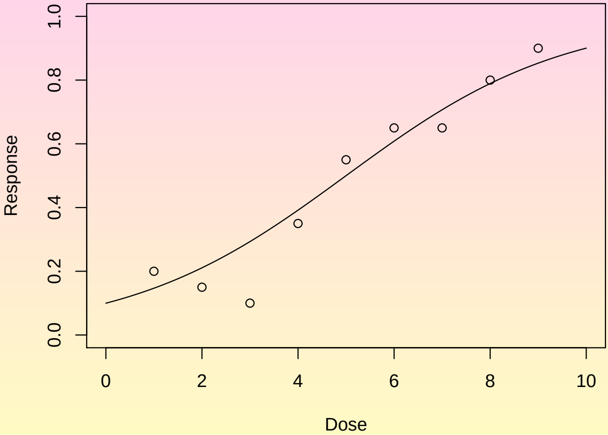
1. Generalisation to bigger problems.
2. It's a simple problem. We're just counting outcomes. It *should* have a simple solution.

Introduction: 1. Generalisations to bigger problems

(i) More than 2 proportions

- Perhaps many.
- Unlikely that we would want independent priors.
- E.g. bioassay.

Introduction: 1. Generalisations to bigger problems – bioassay



Introduction: 1. Generalisations to bigger problems – bioassay

- Y_i out of n_i respond with dose x_i .
- May be many different doses.
- We might not want to use a simple parametric model. We might prefer a nonparametric regression.
- In a design problem there may be many *potential* doses (design points).

Introduction: 1. Generalisations to bigger problems

(ii) More than 2 outcomes.

- Collection of multinomial outcomes rather than binomial outcomes.
- E.g.
  - More complicated contingency tables.
  - Transition matrix in a Markov chain.
  - Item response questionnaires
    - May be ordered categories – eg. student feedback.

Introduction: 2. It ought to be simple.

Single proportion, beta prior — beautifully simple.

- Prior: $\theta \sim \mathrm{beta}(a, b)$
- Posterior: $\theta \sim \mathrm{beta}(a + y, b + n - y)$

Introduction: 2. It ought to be simple.

Single proportion, beta prior — beautifully simple.

- Prior: $\theta \sim \mathrm{beta}(a, b)$
- Posterior: $\theta \sim \mathrm{beta}(a + y, b + n - y)$
- Note: exchangeability of Bernoulli trials $\Rightarrow$ sufficiency of two counts $y, \ (n - y)$

Introduction: 2. It ought to be simple.

Single proportion, beta prior — beautifully simple.

- Prior: $\theta \sim \mathrm{beta}(a, b)$
- Posterior: $\theta \sim \mathrm{beta}(a + y, b + n - y)$
- Note: exchangeability of Bernoulli trials $\Rightarrow$ sufficiency of two counts $y,\ (n - y)$
- Can we not find a simple generalisation?

Introduction: 2. It ought to be simple.

Single proportion, beta prior — beautifully simple.

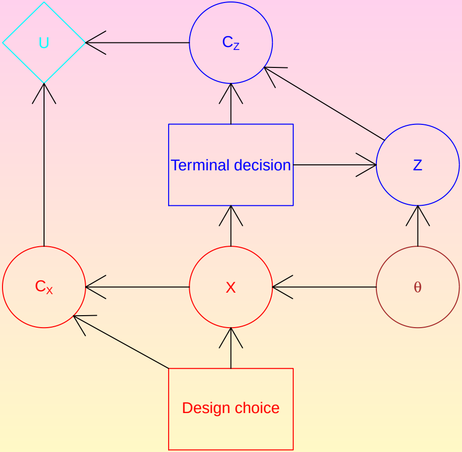
- Prior: $\theta \sim \mathrm{beta}(a, b)$
- Posterior: $\theta \sim \mathrm{beta}(a + y, b + n - y)$
- Note: exchangeability of Bernoulli trials $\Rightarrow$ sufficiency of two counts $y,\ (n - y)$
- Can we not find a simple generalisation?
- . . . or, at least, one that is reasonable tractable?

Introduction: 2. It ought to be simple.

Single proportion, beta prior — beautifully simple.

- Prior: $\theta \sim \mathrm{beta}(a, b)$
- Posterior: $\theta \sim \mathrm{beta}(a + y, b + n - y)$
- Note: exchangeability of Bernoulli trials $\Rightarrow$ sufficiency of two counts $y,\ (n - y)$
- Can we not find a simple generalisation?
- ...or, at least, one that is reasonable tractable?
- Can we have meaningful prior elicitation?

Simple design illustration



Simple design illustration

- Single proportion

Simple design illustration

- Single proportion
- Prior: $\theta \sim \text{beta}(a, b)$

Simple design illustration

- Single proportion
- Prior: $\theta \sim \mathrm{beta}(a, b)$
- We will observe n trials

Simple design illustration

- Single proportion
- Prior: $\theta \sim \mathrm{beta}(a, b)$
- We will observe n trials
- ... then make a terminal decision.

Simple design illustration

- Single proportion
- Prior: $\theta \sim \mathrm{beta}(a, b)$
- We will observe n trials
- ... then make a terminal decision.
- Choose n.

Simple design illustration – Terminal decision

The terminal decision could be many things.

- E.g. Predict the number Z of successes out of m future trials.
- Introduce a benefit utility

$$U_{b,n}(Z, P)$$

- E.g.

Simple design illustration – Terminal decision

The terminal decision could be many things.

- E.g. Predict the number Z of successes out of m future trials.
- Introduce a benefit utility

$$U_{b,n}(Z, P)$$

- E.g.

  (i)

$$U_{b,n}(Z, P) = 1 - \frac{|Z - P|}{m}$$

Simple design illustration – Terminal decision

The terminal decision could be many things.

- E.g. Predict the number Z of successes out of m future trials.
- Introduce a benefit utility

$$U_{b,n}(Z, P)$$

- E.g.

(i)

$$U_{b,n}(Z, P) = 1 - \frac{|Z - P|}{m}$$

(ii)

$$U_{b,n}(Z, P) = 1 - \left(\frac{Z - P}{m}\right)^2$$

Simple design illustration – Terminal decision

The terminal decision could be many things.

- E.g. Predict the number Z of successes out of m future trials.
- Introduce a benefit utility

$$U_{b,n}(Z, P)$$

- E.g.
  (i)
  $$U_{b,n}(Z, P) = 1 - \frac{|Z - P|}{m}$$
  (ii)
  $$U_{b,n}(Z, P) = 1 - \left(\frac{Z - P}{m}\right)^2$$
  (iii) We get to choose P. This is the terminal decision.

Simple design illustration – Terminal decision

$$
\begin{aligned}
U_{b,n}(Z,P) &= 1 - \left(\frac{Z-P}{m}\right)^2 \\
&= 1 - \frac{1}{m^2}\{Z - \mathrm{E}_{Z|\theta}(Z \mid \theta) + \mathrm{E}_{Z|\theta}(Z \mid \theta) - P\}^2 \\
&= 1 - \frac{1}{m^2}\{[Z - \mathrm{E}_{Z|\theta}(Z \mid \theta)]^2 + [\mathrm{E}_{Z|\theta}(Z \mid \theta) - P]^2 \\
&\quad +2[Z - E_{Z|\theta}(Z \mid \theta)][E_{Z|\theta}(Z \mid \theta) - P]\}
\end{aligned}
$$

Simple design illustration – Terminal decision

$$U_{b,n}(Z,P) \;=\; 1 - \frac{1}{m^2}\{[Z - \mathrm{E}_{Z|\theta}(Z \mid \theta)]^2 + [\mathrm{E}_{Z|\theta}(Z \mid \theta) - P]^2 + 2[Z - E_{Z|\theta}(Z \mid \theta)][E_{Z|\theta}(Z \mid \theta) - P]\}$$

Take expectations over $Z \mid \theta$:

$$\begin{aligned}\mathrm{E}_{Z|\theta}[U_{b,n}(Z,P)] &= 1 - \frac{1}{m^2}\mathrm{Var}_{Z|\theta}(Z \mid \theta) - \frac{1}{m^2}[\mathrm{E}_{Z|\theta}(Z \mid \theta) - P]^2 \\ &= 1 - \frac{1}{m^2} m\theta(1-\theta) - \frac{1}{m^2}(m\theta - P)^2\end{aligned}$$

Simple design illustration – Terminal decision

$$
\begin{aligned}
\mathrm{E}_{Z|\theta}[U_{b,n}(Z,P)] &= 1 - \frac{1}{m^2} m\theta(1-\theta) - \frac{1}{m^2}(m\theta - P)^2 \\
&= 1 - \frac{1}{m^2} m\theta(1-\theta) \\
&\quad - \frac{1}{m^2}(m\theta - m\mathrm{E}_\theta(\theta) + m\mathrm{E}_\theta(\theta) - P)^2 \\
&= 1 - \frac{1}{m^2} m\theta(1-\theta) - [\theta - \mathrm{E}_\theta(\theta)]^2 \\
&\quad - [\mathrm{E}_\theta(\theta) - P/m]^2 \\
&\quad + 2[\theta - \mathrm{E}_\theta(\theta)][\mathrm{E}_\theta(\theta) - P/m]
\end{aligned}
$$

Simple design illustration – Terminal decision

$$\begin{aligned}
\mathrm{E}_{Z|\theta}[U_{b,n}(Z,P)] &= 1-\frac{1}{m^2}m\theta(1-\theta)-[\theta-\mathrm{E}_\theta(\theta)]^2 \\
&\quad -[\mathrm{E}_\theta(\theta)-P/m]^2 \\
&\quad +2[\theta-\mathrm{E}_\theta(\theta)][\mathrm{E}_\theta(\theta)-P/m]
\end{aligned}$$

Take expectations over θ:

$$\mathrm{E}_{Z,\theta}[U_{b,n}(Z,P)] = 1-\frac{1}{m}\mathrm{E}_\theta[\theta(1-\theta)]-\mathrm{Var}_\theta(\theta)-[\mathrm{E}_\theta(\theta)-P/m]^2$$

Simple design illustration – Terminal decision

$$\mathrm{E}_{Z,\theta}[U_{b,n}(Z,P)] = 1 - \frac{1}{m}\mathrm{E}_\theta[\theta(1-\theta)] - \mathrm{Var}_\theta(\theta) - [\mathrm{E}_\theta(\theta) - P/m]^2$$

Maximise this expectation by setting $P/m = \mathrm{E}_\theta(\theta)$.
That is $P = m\mathrm{E}_\theta(\theta)$.

Simple design illustration – Terminal decision

We choose $P = m\mathrm{E}_\theta(\theta)$. After our experiment, when we have observed n trials with x successes, we have

$$\mathrm{E}_\theta(\theta) = \mathrm{E}_1(\theta \mid x) = \hat{\theta} = \frac{a+x}{a+b+n}$$

so we choose

$$\frac{P}{m} = \frac{a+x}{a+b+n}.$$

Note the use of the explicit formula.

Simple design illustration – Choice of sample size

So, our utility *before* the experiment is

$$
\begin{aligned}
U^*_{b,n}(\theta,\hat{\theta}) &= 1 - \frac{\theta(1-\theta)}{m} - (\theta-\hat{\theta})^2 \\
&= 1 - \frac{\theta(1-\theta)}{m} - \left(\theta - \frac{a+x}{a+b+n}\right)^2 \\
&= 1 - \frac{\theta(1-\theta)}{m} - \left[\frac{(a+b)\theta - a}{a+b+n}\right]^2 \\
&\quad - \left[\frac{x-n\theta}{a+b+n}\right]^2 + 2\frac{(x-n\theta)[(a+b)\theta-a]}{(a+b+n)^2}
\end{aligned}
$$

(after some algebra).

Simple design illustration – Choice of sample size

$$U^*_{b,n}(\theta,\hat{\theta}) = 1 - \frac{\theta(1-\theta)}{m} - \left[\frac{(a+b)\theta - a}{a+b+n}\right]^2 - \left[\frac{x-n\theta}{a+b+n}\right]^2 + 2\frac{(x-n\theta)[(a+b)\theta - a]}{(a+b+n)^2}$$

Take expectations over $X \mid \theta$.

$$\mathrm{E}_{X|\theta}[U^*_{b,n}(\theta,\hat{\theta})] = 1 - \frac{\theta(1-\theta)}{m} - \left\{\frac{(a+b)(\theta - a/[a+b])}{a+b+n}\right\}^2 - \frac{n\theta(1-\theta)}{(a+b+n)^2}$$

(again, after some algebra).

Simple design illustration – Choice of sample size

Finally we take expectations over the *prior* distribution of θ. After some algebra (again):

$$\mathrm{E}(U^*_{b,n}) = 1 - \frac{ab}{(a+b)(a+b+1)} \left\{ \frac{1}{m} + \frac{1}{a+b+n} \right\}$$

Simple design illustration – Choice of sample size

Finally we take expectations over the *prior* distribution of θ.
After some algebra (again):

$$\mathrm{E}(U^*_{b,n}) = 1 - \frac{ab}{(a+b)(a+b+1)} \left\{ \frac{1}{m} + \frac{1}{a+b+n} \right\}$$

- So we have an explicit formula which we can combine with a cost utility and then we can maximise the result wrt n ...

Simple design illustration – Choice of sample size

Finally we take expectations over the *prior* distribution of θ.
After some algebra (again):

$$\mathrm{E}(U^*_{b,n}) = 1 - \frac{ab}{(a+b)(a+b+1)}\left\{\frac{1}{m} + \frac{1}{a+b+n}\right\}$$

- So we have an explicit formula which we can combine with a cost utility and then we can maximise the result wrt n ...
- ... but notice how this depends on having explicit formulae for the necessary expectations and for the terminal decision rule.

Measures of association – 2 proportions

- Familiar with bivariate normal distribution – 5 parameters:
  - 2 means
  - 2 variances
  - 1 (product-moment) correlation (or covariance)
- The same approach might not be appropriate for proportions where $0 < \theta < 1$.

Measures of association – 2 proportions

Some possible alternatives:

Measures of association – 2 proportions

Some possible alternatives:

- Rank correlation
    - Kendall's τ
    - Spearman's ρ

Measures of association – 2 proportions

Some possible alternatives:

- Rank correlation
  - Kendall's τ
  - Spearman's ρ
- Pearson (product-moment) correlation applied to transformed unknowns (η_1, η_2)

$$\eta_i = g(\theta_i)$$

$g()$: logits, probits, whatever.

Measures of association – 2 proportions

Some possible alternatives:

- Rank correlation
  - Kendall's τ
  - Spearman's ρ
- Pearson (product-moment) correlation applied to transformed unknowns (η_1, η_2)

$$\eta_i = g(\theta_i)$$

  $g()$: logits, probits, whatever.
- Directly in terms of observables.

Measures of association – Kendall's τ

- Suppose (θ_1, θ_2) have some bivariate distribution.
- Consider observing a sequence of independent draws $(T_{1,j}, T_{2,j}),\ j = 1, 2, 3, \ldots$ from this distribution.
- Kendall's τ:

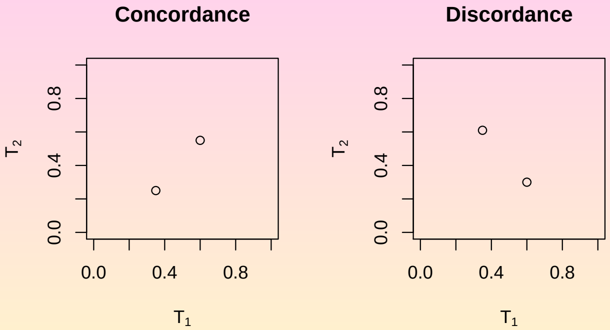
$$\begin{aligned}\tau_{1,2} = \quad & \Pr\{(T_{1,1} - T_{1,2})(T_{2,1} - T_{2,2}) > 0\} \\ - \quad & \Pr\{(T_{1,1} - T_{1,2})(T_{2,1} - T_{2,2}) < 0\}\end{aligned}$$

"Probability of concordance minus probability of discordance"

Equivalently

$$\tau_{1,2} = 2\Pr\{(T_{1,1} - T_{1,2})(T_{2,1} - T_{2,2}) > 0\} - 1$$

Measures of association – Kendall's τ



Measures of association – Spearman's ρ

$$\begin{aligned}\rho_{1,2} = \quad & 3 \quad \Pr\{(T_{1,1} - T_{1,2})(T_{2,1} - T_{2,3}) > 0\} \\ & - \quad \Pr\{(T_{1,1} - T_{1,2})(T_{2,1} - T_{2,3}) < 0\}\end{aligned}$$

- Note: $T_{1,2}$, $T_{2,3}$ independent.
- Interpretation not as straightforward.

Measures of association – Transformations

"Ordinary" product-moment correlation of (η_1, η_2) where $\eta_i = g(\theta_i)$ (logits, probits, whatever).

- Choice of transformation a bit arbitrary.
- Elicitation a little tricky?

Measures of association – Directly from observables

Bernoulli trial j with $\theta = \theta_i$:

$$X_{i,j} = \begin{cases} 1 \\ 0 \end{cases}$$

$$\begin{aligned} \mathrm{E}(X_{i,1}) &= \mathrm{E}(\theta_i) \\ \mathrm{E}(X_{i,1}X_{i,2}) &= \mathrm{E}(\theta_i^2) \\ \mathrm{E}(X_{1,1}X_{2,1}) &= \mathrm{E}(\theta_1\theta_2) \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{Var}(\theta_i) &= \mathrm{E}(X_{i,1}X_{i,2}) - [\mathrm{E}(X_{i,1})]^2 \\ \mathrm{Covar}(\theta_1, \theta_2) &= \mathrm{E}(X_{1,1}X_{2,1}) - \mathrm{E}(X_{1,1})\mathrm{E}(X_{2,1}) \end{aligned}$$

Measures of association – Directly from observables

This seems to argue in favour of simple product-moment covariance/correlation for (θ_1, θ_2) but ...

Measures of association – Directly from observables

This seems to argue in favour of simple product-moment covariance/correlation for (θ_1, θ_2) but ...

- would such elicitation work in practice (bearing in mind the mean-variance relationship)?

Measures of association – Directly from observables

This seems to argue in favour of simple product-moment covariance/correlation for (θ_1, θ_2) but ...

- would such elicitation work in practice (bearing in mind the mean-variance relationship)?
- can we relate these moments to parameters of tractable joint distributions?

Some possible joint belief structures

Some possible joint belief structures

1. Full probability distributions.

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions
   (iii) Copulas

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions
   (iii) Copulas
   (iv) Multipliers — and mixtures

Some possible joint belief structures

1. Full probability distributions.
   - (i) Dirichlet distribution — and mixtures
   - (ii) Hierarchical beta distributions
   - (iii) Copulas
   - (iv) Multipliers — and mixtures
2. Partial belief specification

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions
   (iii) Copulas
   (iv) Multipliers — and mixtures
2. Partial belief specification
   (i) Requirements

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions
   (iii) Copulas
   (iv) Multipliers — and mixtures
2. Partial belief specification
   (i) Requirements
   (ii) Bayes linear kinematics — (*etc*).

Some possible joint belief structures

1. Full probability distributions.
   (i) Dirichlet distribution — and mixtures
   (ii) Hierarchical beta distributions
   (iii) Copulas
   (iv) Multipliers — and mixtures
2. Partial belief specification
   (i) Requirements
   (ii) Bayes linear kinematics — (*etc*).
   (iii) Direct counts method

Dirichlet distribution

Consider again 2×2 table.

	Outcome 0	Outcome 1
Treatment 1	$n_1 - Y_1$	Y_1
Treatment 2	$n_2 - Y_2$	Y_2

Imagine a population of individuals with four types, as follows.

Outcomes		
Treatment 1	Treatment 2	Probability
1	1	π_{11}
1	0	π_{10}
0	1	π_{01}
0	0	π_{00}

- Let $\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00} \sim \text{Dirichlet}(a_{11}, a_{10}, a_{01}, a_{00})$.
- 4 parameters — better than 2 but not quite enough.

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$
- Likelihood proportional to
  $(\pi_{11} + \pi_{10})^{y_1}(\pi_{01} + \pi_{00})^{n_1-y_1}(\pi_{11} + \pi_{01})^{y_2}(\pi_{10} + \pi_{00})^{n_2-y_2}$

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$
- Likelihood proportional to
  $(\pi_{11} + \pi_{10})^{y_1}(\pi_{01} + \pi_{00})^{n_1-y_1}(\pi_{11} + \pi_{01})^{y_2}(\pi_{10} + \pi_{00})^{n_2-y_2}$
- Hence the posterior is a finite mixture of Dirichlet distributions . . .

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$
- Likelihood proportional to $(\pi_{11}+\pi_{10})^{y_1}(\pi_{01}+\pi_{00})^{n_1-y_1}(\pi_{11}+\pi_{01})^{y_2}(\pi_{10}+\pi_{00})^{n_2-y_2}$
- Hence the posterior is a finite mixture of Dirichlet distributions . . .
- . . . so why not start with a prior which is a mixture of Dirichlet distributions? — Extra parameter(s).

Dirichlet distribution

$$\theta_1 = \pi_{11} + \pi_{10} \qquad 1 - \theta_1 = \pi_{01} + \pi_{00}$$
$$\theta_2 = \pi_{11} + \pi_{01} \qquad 1 - \theta_2 = \pi_{10} + \pi_{00}$$

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$
- Likelihood proportional to $(\pi_{11} + \pi_{10})^{y_1}(\pi_{01} + \pi_{00})^{n_1-y_1}(\pi_{11} + \pi_{01})^{y_2}(\pi_{10} + \pi_{00})^{n_2-y_2}$
- Hence the posterior is a finite mixture of Dirichlet distributions . . .
- . . . so why not start with a prior which is a mixture of Dirichlet distributions? — Extra parameter(s).
- Need a suitable family of mixtures.
    - Various possibilities.

Dirichlet distribution – mixtures

The Dirichlet parameters can (sort of) be seen as counts of prior observations of the four types of individual. Suppose we introduce four more types of “prior individual” with “prior counts” $b_{1,*}, b_{0,*}, b_{*,1}, b_{*,0}$.

Imagine a population of individuals with four types, as follows.

Outcomes		
Treatment 1	Treatment 2	Probability
1	?	$\pi_{11} + \pi_{10}$
0	?	$\pi_{01} + \pi_{00}$
?	1	$\pi_{11} + \pi_{01}$
?	0	$\pi_{10} + \pi_{00}$

Dirichlet distribution – mixtures

Dirichlet distribution – mixtures

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$

  $\times(\pi_{11}+\pi_{10})^{b_{1*}}(\pi_{01}+\pi_{00})^{b_{0*}}(\pi_{11}+\pi_{01})^{b_{*1}}(\pi_{10}+\pi_{00})^{b_{*0}}$

Dirichlet distribution – mixtures

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$

  $\times(\pi_{11}+\pi_{10})^{b_{1*}}(\pi_{01}+\pi_{00})^{b_{0*}}(\pi_{11}+\pi_{01})^{b_{*1}}(\pi_{10}+\pi_{00})^{b_{*0}}$

- Recall that the likelihood is proportional to $(\pi_{11}+\pi_{10})^{y_1}(\pi_{01}+\pi_{00})^{n_1-y_1}(\pi_{11}+\pi_{01})^{y_2}(\pi_{10}+\pi_{00})^{n_2-y_2}$

Dirichlet distribution – mixtures

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$

  $\times(\pi_{11}+\pi_{10})^{b_{1*}}(\pi_{01}+\pi_{00})^{b_{0*}}(\pi_{11}+\pi_{01})^{b_{*1}}(\pi_{10}+\pi_{00})^{b_{*0}}$

- Recall that the likelihood is proportional to $(\pi_{11}+\pi_{10})^{y_1}(\pi_{01}+\pi_{00})^{n_1-y_1}(\pi_{11}+\pi_{01})^{y_2}(\pi_{10}+\pi_{00})^{n_2-y_2}$
- We now have 8 parameters — more than enough!

Dirichlet distribution – mixtures

- Prior density proportional to $\pi_{11}^{a_{11}-1}\pi_{10}^{a_{10}-1}\pi_{01}^{a_{01}-1}\pi_{00}^{a_{00}-1}$

  $$\times(\pi_{11}+\pi_{10})^{b_{1*}}(\pi_{01}+\pi_{00})^{b_{0*}}(\pi_{11}+\pi_{01})^{b_{*1}}(\pi_{10}+\pi_{00})^{b_{*0}}$$

- Recall that the likelihood is proportional to $(\pi_{11}+\pi_{10})^{y_1}(\pi_{01}+\pi_{00})^{n_1-y_1}(\pi_{11}+\pi_{01})^{y_2}(\pi_{10}+\pi_{00})^{n_2-y_2}$
- We now have 8 parameters — more than enough!
- ... but can we really elicit 8 parameters for just two proportions?

Hierarchical beta distribtuions

Based on a suggestion by Sarah Germain.

$$
\begin{aligned}
\theta_i \mid \mu &\sim \mathrm{beta}(k\mu,\ k[1-\mu]) \\
\mu &\sim \mathrm{beta}(c, d)
\end{aligned}
$$

Joint density

$$
\begin{aligned}
f(\mu, \theta_1, \theta_2) &= B^{-1}(c, d)\mu^{c-1}(1-\mu)^{d-1} \\
&\quad \times B^{-1}(k\mu, k(1-\mu))\theta_1^{k\mu-1}(1-\theta_1)^{k(1-\mu)-1} \\
&\quad \times B^{-1}(k\mu, k(1-\mu))\theta_2^{k\mu-1}(1-\theta_2)^{k(1-\mu)-1}
\end{aligned}
$$

Hierarchical beta distribtuions

Generalise to:

$$
\begin{aligned}
\theta_i \mid \mu &\sim \mathrm{beta}(k_i\mu + a_i,\ k_i[1-\mu] + b_i) \\
\mu &\sim \mathrm{beta}(c, d) \\
f(\mu, \theta_1, \theta_2) &= B^{-1}(c, d)\mu^{c-1}(1-\mu)^{d-1} \\
&\quad \times B^{-1}(k_1\mu + a_1, k_1(1-\mu) + b_1) \\
&\quad \times \theta_1^{k_1\mu + a_1 - 1}(1-\theta_1)^{k_1(1-\mu)+b_1-1} \\
&\quad \times B^{-1}(k_2\mu + a_2, k_2(1-\mu) + b_2) \\
&\quad \times \theta_2^{k_2\mu + a_2 - 1}(1-\theta_2)^{k_2(1-\mu)+b_2-1}
\end{aligned}
$$

Hierarchical beta distribtuions

- Now conjugate . . .

Hierarchical beta distribtuions

- Now conjugate . . .
- . . . but marginal for θ_i not very nice.

Hierarchical beta distribtuions

- Now conjugate . . .
- . . . but marginal for θ_i not very nice.
- Needs more work to make the correlation structure flexible.

Hierarchical beta distribtuions

- Now conjugate . . .
- . . . but marginal for θ_i not very nice.
- Needs more work to make the correlation structure flexible.
- Still, could be useful for some problems.

Copulas

Suppose, for example,

$$\begin{aligned} \theta_i &\sim \mathrm{beta}(a_i, b_i) \\ U_i &= G_i(\theta_i) \end{aligned}$$

where $G_i()$ is the cdf of $\text{beta}(a_i, b_i)$. Hence

$$U_i \sim U(0, 1)$$

Introduce $H(u_1, u_2)$ as a joint cdf for U_1, U_2.

Copulas

Suppose, for example,

$$\begin{aligned} \theta_i &\sim \mathrm{beta}(a_i, b_i) \\ U_i &= G_i(\theta_i) \end{aligned}$$

where $G_i()$ is the cdf of $\text{beta}(a_i, b_i)$. Hence

$$U_i \sim U(0,1)$$

Introduce $H(u_1, u_2)$ as a joint cdf for U_1, U_2.

- If $H(,)$ is a *copula* then it preserves the uniform marginal distributions of U_1, U_2 but makes them dependent. Hence the beta marginal distributions of θ_1, θ_2 are also preserved.

- Joint pdf of (θ_1, θ_2) is

$$\begin{aligned} f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1 \partial\theta_2} H(u_1, u_2) \\ &= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\ &= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2)) \end{aligned}$$

where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.

Copulas

- Joint pdf of (θ_1, θ_2) is

$$
\begin{aligned}
f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1 \partial\theta_2} H(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2))
\end{aligned}
$$

  where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.
- Useful for *marginal* elicitation.

Copulas

- Joint pdf of (θ_1, θ_2) is

$$\begin{aligned} f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1 \partial\theta_2} H(u_1, u_2) \\ &= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\ &= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2)) \end{aligned}$$

  where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.
- Useful for *marginal* elicitation.
- Rank correlation just depends on the choice of copula.

Copulas

- Joint pdf of (θ_1, θ_2) is

$$
\begin{aligned}
f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1\partial\theta_2} H(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2))
\end{aligned}
$$

  where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.
- Useful for *marginal* elicitation.
- Rank correlation just depends on the choice of copula.
- But standard copula families give limited range of correlation.

Copulas

- Joint pdf of (θ_1, θ_2) is

$$
\begin{aligned}
f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1 \partial\theta_2} H(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2))
\end{aligned}
$$

  where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.
- Useful for *marginal* elicitation.
- Rank correlation just depends on the choice of copula.
- But standard copula families give limited range of correlation.
- Prior is conjugate

Copulas

- Joint pdf of (θ_1, θ_2) is

$$
\begin{aligned}
f(\theta_1, \theta_2) &= \frac{\partial^2}{\partial\theta_1\partial\theta_2} H(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(u_1, u_2) \\
&= g_1(\theta_1) g_2(\theta_2) h(G(\theta_1), G(\theta_2))
\end{aligned}
$$

  where $h(,)$ is the *copula density* and $g()$ is (eg) the beta density.
- Useful for *marginal* elicitation.
- Rank correlation just depends on the choice of copula.
- But standard copula families give limited range of correlation.
- Prior is conjugate
- But posterior is no longer a copula so marginals are not nice.

Multipliers — and mixtures

Recall copula. Joint density:

$$g_1(\theta_1)g_2(\theta_2)h(G(\theta_1), G(\theta_2))$$

Multipliers — and mixtures

Recall copula. Joint density:

$$g_1(\theta_1)g_2(\theta_2)h(G(\theta_1), G(\theta_2))$$

- What if $h(,)$ is *not* a copula density?

Multipliers — and mixtures

Recall copula. Joint density:

$$g_1(\theta_1)g_2(\theta_2)h(G(\theta_1), G(\theta_2))$$

- What if $h(,)$ is *not* a copula density?
- We could make it something convenient like a polynomial in (θ_1, θ_2).

Multipliers — and mixtures

Recall copula. Joint density:

$$g_1(\theta_1)g_2(\theta_2)h(G(\theta_1), G(\theta_2))$$

- What if $h(,)$ is *not* a copula density?
- We could make it something convenient like a polynomial in (θ_1, θ_2).
- Joint density proportional to

$$g_1(\theta_1)g_2(\theta_2)h(\theta_1, \theta_2)$$

Multipliers — and mixtures

Recall copula. Joint density:

$$g_1(\theta_1)g_2(\theta_2)h(G(\theta_1), G(\theta_2))$$

- What if $h(,)$ is *not* a copula density?
- We could make it something convenient like a polynomial in (θ_1, θ_2).
- Joint density proportional to

$$g_1(\theta_1)g_2(\theta_2)h(\theta_1, \theta_2)$$

- If $h(,)$ is a polynomial we get a mixture density

$$f(\theta_1, \theta_2) = \sum \pi_k g_{1,k}(\theta_1)g_{2,k}(\theta_2)$$

where each $g_{i,k}$ is a beta density.

Multipliers — and mixtures

- Eg (for positive correlation)

$$h(\theta_1, \theta_2) = [(1 + \theta_1 - \theta_2)(1 - \theta_1 + \theta_2)]^q$$

with $q > 0$.

Multipliers — and mixtures

- Eg (for positive correlation)

$$h(\theta_1, \theta_2) = [(1 + \theta_1 - \theta_2)(1 - \theta_1 + \theta_2)]^q$$

  with $q > 0$.

- $h(,)$ is maximised when $\theta_1 = \theta_2$

Multipliers — and mixtures

- Eg (for positive correlation)

$$h(\theta_1, \theta_2) = [(1 + \theta_1 - \theta_2)(1 - \theta_1 + \theta_2)]^q$$

  with $q > 0$.
- $h(,)$ is maximised when $\theta_1 = \theta_2$
- ... but this might not be what we want. We might want, for example, a tendency towards $\theta_1 = \theta_2^c$.

Multipliers — and mixtures

- Eg (for positive correlation)

$$h(\theta_1, \theta_2) = [(1 + \theta_1 - \theta_2)(1 - \theta_1 + \theta_2)]^q$$

  with $q > 0$.
- $h(,)$ is maximised when $\theta_1 = \theta_2$
- ... but this might not be what we want. We might want, for example, a tendency towards $\theta_1 = \theta_2^c$.
- Also, of course, it makes the marginals a bit more complicated — elicitation.

Multipliers — and mixtures

- Eg (for negative correlation)

$$h(\theta_1, \theta_2) = [(\theta_1 + \theta_2)(2 - \theta_1 - \theta_2)]^q$$

with $q > 0$.

Multipliers — and mixtures

- Eg (for negative correlation)

$$h(\theta_1, \theta_2) = [(\theta_1 + \theta_2)(2 - \theta_1 - \theta_2)]^q$$

  with $q > 0$.
- $h(,)$ is maximised when $\theta_1 + \theta_2 = 1$

Multipliers — and mixtures

- Eg (for negative correlation)

$$h(\theta_1, \theta_2) = [(\theta_1 + \theta_2)(2 - \theta_1 - \theta_2)]^q$$

  with $q > 0$.
- $h(,)$ is maximised when $\theta_1 + \theta_2 = 1$
- Comments as for previous example.

Multipliers — and mixtures

- Eg (for negative correlation)

$$h(\theta_1, \theta_2) = [(\theta_1 + \theta_2)(2 - \theta_1 - \theta_2)]^q$$

  with $q > 0$.

- $h(,)$ is maximised when $\theta_1 + \theta_2 = 1$
- Comments as for previous example.
- Further work needed . . .