Intrinsic meta-analysis of contingency tables

George Casella\textsuperscript{1,*;†;§} and Elías Moreno\textsuperscript{2;‡;¶}

\textsuperscript{1}Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.
\textsuperscript{2}Department of Statistics, University of Granada, 18071, Granada, Spain

SUMMARY

Meta-analysis has a natural formulation as a Bayesian hierarchical model. The main theoretical difficulty is the construction of a sensible relationship between the parameters of the individual statistical experiments and the meta-parameter. Since that prior information on such a relationship is typically not available, we argue that this relationship should be dictated by the structure of the model at hand. We then propose a novel procedure based on intrinsic priors which we fully develop for the case of meta-analysis of $2 \times 2$ contingency tables. Illustrations on real and artificial tables are given. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: intrinsic priors; Monte Carlo integration; chi-squared tests

1. INTRODUCTION

Statistical methods for pooling the findings of a collection of statistical analyses is called a meta-analysis. For example, a clinical trial for establishing the effectiveness of a new treatment may be carried out in $m$ different hospitals. The quantity of interest is often the effectiveness of the new treatment, a meta-parameter which is non-observable in nature, and it has to be derived from the performance of the new treatment in the hospitals for which data are available.

\textsuperscript{*}Correspondence to: George Casella, Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.
\textsuperscript{†}E-mail: casella@stat.ufl.edu
\textsuperscript{‡}E-mail: emoreno@ugr.es
\textsuperscript{§}Distinguished Professor and Chair. This work was done while Casella was on sabbatical at the University of Granada.
\textsuperscript{¶}Professor.

Contract/grant sponsor: National Science Foundation; contract/grant number: DMS-9971586
Contract/grant sponsor: Ministerio de Ciencia y Tecnología; contract/grant number: BEC2001-2982

Copyright © 2005 John Wiley & Sons, Ltd.  Accepted July 2004
A possible formulation of the above situation is as follows. Assume that there are available $m$ statistical experiments with sampling distributions

$$\{f(x_k|\theta), \quad k = 1, \ldots, m\}$$

where, in experiment $k$, $x_k$ and $\theta$ are the observed data and parameters, respectively. Now, let $f(x|\theta)$ represent a meta-model, a statistical model for which neither $x$ nor $\theta$ are observable, $f$ being of the same form as the sampling distributions of the experiments. The goal is to do inference on the meta-parameter $\theta$ based on the entire data set $\{x_k, \quad k = 1, \ldots, m\}$. We also note that parameters $p(k)$ and $\theta$ are typically multidimensional. For instance, in $2 \times 2$ contingency tables with fixed totals, $p(k)$ represents the parameter of a multinomial sampling distribution with four cells and the meta-parameter $\theta$ represents the multinomial parameter of a $2 \times 2$ meta-contingency table.

A difficulty in meta-analysis is in the establishment of a sensible relationship between $p(k)$ and $\theta$. In the literature this relationship has been derived through either an asymptotic argument (for instance, see the analyses on the log odds ratio in contingency tables based on an asymptotic chi-squared distribution by Yusuf et al. [1] and DerSimonian and Laird [2]), or a hierarchical model in which a transformation of $p(k)$ is normally distributed around $\theta$ (for instance, the analysis on the log odds in contingency table by Carlin [3], and Shi and Copas [4]). The distribution of the location and scale hyperparameters of the normal distribution is assumed to be the conventional default improper prior (see also Reference [5]). The above analyses mainly focus on estimating parameters rather than on testing problems.

But, if we are to proceed in the spirit of a meta-analysis, with the goal of making a policy decision based on the meta-parameter, then estimation of the effect sizes in each experiment is tangential. Although inferences on the individual $p(k)$ might also be of interest, in a meta-analysis there is no need for doing the statistical analysis of the individual experiments. For example, if the experiment is to decide whether a new treatment is better than the standard, the decision should be based on the meta-parameter. If certain hospitals have higher success rates than others, this makes no difference in the meta-conclusion. (We are not saying that estimation of individual effects is not important, just that it has no bearing on the meta-conclusion.) Thus, here we will only concentrate on testing and estimation of the meta-parameter.

Instead of basing the analysis on the usual assumption that the estimator of $p(k)$ (or some convenient function of it) is normally distributed, so making more plausible the normal distribution for $p(k)$ centred at $\theta$, we generate the distribution of $p(k)$ conditional on $\theta$ by essentially using the structure of the model $f(x|\theta)$. More precisely, we propose as the distribution for $p(k)$, conditional on $\theta$, the intrinsic prior arising from the comparison of model $f(x|\theta)$ with $\{f(x_k|\theta), \pi^D(p(k))\}$, where $\theta$ is an unknown but fixed point and $\pi^D$ is the conventional default prior for $p(k)$.

In comparing the meta-model $f(x|\theta)$ with a Bayesian default experiment $\{f(x_k|\theta), \pi^D(p(k))\}$, we observe that $f(x|\theta)$ is nested in $\{f(x_k|p(k)), p(k) \in \Theta\}, \Theta$ being the parameter space of any of the parameters considered. This allows us to use the standard theory of the intrinsic prior [6–8]. Hence, the intrinsic prior for $p(k)$ conditional on $\theta$ is given by

$$\pi^I(p(k)|\theta) = \pi^D(p(k))E_{x|p(k)}f(x|\theta) m_k(x)$$

(2)
where \( m_k(x) = \int_\Theta f(x|p^{(k)}) \pi^D(p^{(k)}) \, dp^{(k)} \). We remark that \( \pi^I(p^{(k)}|\theta) \) is a proper prior, that is \( \int_\Theta \pi^I(p^{(k)}|\theta) \, dp^{(k)} = 1 \), regardless whether or not the default prior \( \pi^D(p^{(k)}) \) is proper. In the case where \( \pi^D(p^{(k)}) \) is improper, one simply needs to consider \( x \) as a vector of minimal size so that \( m_k(x) \) is finite.

Note that we do not need to assume that the individual densities in (1) are identical; the \( p^{(k)} \) can be different, and contain covariates. Indeed, in our examples the individual experiments have tables with different totals, which are covariates. So, given the data \( x_k \), the induced likelihood on the meta-parameter \( \theta \) from the \( k \)th experiment and the intrinsic prior in (2) is obtained as

\[
f(x_k|\theta) = \int_\Theta f(x_k|p^{(k)}) \pi^I(p^{(k)}|\theta) \, dp^{(k)}
\]

In what follows we will assume that the data \( x = \{x_k, \ k = 1, \ldots, m\} \) are, conditional on \( \theta \), independent. That is, the joint density \( f(x|\theta) \) is given by

\[
f(x|\theta) = \prod_k f(x_k|\theta)
\]

A default Bayesian analysis on the meta-parameter is now possible by using the density \( f(x|\theta) \) and its corresponding default prior \( \pi^D(\theta) \). For example, posterior distributions can be calculated and compared, and the meta-parameter can be estimated. This is the approach taken in Section 3.

For inference on the parameter \( p^{(k)} \) of an individual experiment, conditional on the entire data set \( x \), the posterior density of \( p^{(k)} \) can be obtained from

\[
\pi(p^{(k)}|x) = \int_\Theta \pi^I(p^{(k)}|\theta) \pi(\theta|x) \, d\theta
\]

where

\[
\pi(\theta|x) = \frac{f(x|\theta) \pi^D(\theta)}{\int_\Theta f(x|\theta) \pi^D(\theta) \, d\theta}
\]

If desired, we can test \( H_0: \theta = \theta_0 \) versus \( H_1: \theta \neq \theta_0 \). Assuming that a priori \( P(H_0) = P(H_1) = 1/2 \), the posterior probability of \( H_0 \) conditional on the data set \( x = \{x_k, \ k = 1, \ldots, m\} \) is obtained as

\[
P(H_0|x) = \frac{1}{1 + B_{10}(x)}
\]

where the Bayes factor \( B_{10} \) is

\[
B_{10}(x) = \int f(x|\theta) \pi^I(\theta|\theta_0) \, d\theta
\]

with \( f(x|\theta) \) given in (4) and \( \pi^I(\theta|\theta_0) \) the intrinsic prior for \( \theta \) conditional on \( \theta_0 \).

In a similar way a composite null hypothesis can also be tested, hypotheses that can arise in testing the effectiveness of a treatment. For example, in testing the equality of two treatments or, equivalently, testing independence of the meta-parameter using a multinomial model,
a default Bayesian test is between the models.

\[ M_0 : \{ f(z|\theta_0, n), \quad \pi^D(\theta, \xi) = I_{(0,1)}(\theta)I_{(0,1)}(\xi) \} \]

and

\[ M_1 : \{ f(z|\theta, n), \quad \pi^D(\theta) = \text{Dirichlet}(\theta|1, 1, 1, 1) \} \]

where \( f(z|\theta, n) \) is the multinomial pmf with parameters \( \theta \) and \( n \), and \( I_d \) denotes the indicator function of the set \( A \). We test

\[ H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0 \]

where

\[ \theta_0 = \begin{pmatrix} \theta \xi \\ (1 - \theta) \xi \\ \theta (1 - \xi) \\ (1 - \theta)(1 - \xi) \end{pmatrix} \]

However, for testing independence in contingency tables Casella and Moreno [9] argue that the Dirichlet prior distribution is not appropriate since models in the alternative far from the null are considered as equally plausible as those that are closer to the null (see also References [10–12]). They recommend replacing the Dirichlet prior in \( M_1 \) with the intrinsic prior for \( \theta \), conditional on \( H_0 \). From a comparative analysis with the volume test by Diaconis and Efron [13], the hierarchical formulation with Dirichlet-log Cauchy distribution by Good and Crook [14], and the Dirichlet prior itself, Casella and Moreno [9] show that the performance of the test based on the intrinsic prior considerably improves on the others.

We note that this method does not need any assumption on the distribution of any ‘suitable’ statistics, but the resulting distribution of \( p^{(k)} \), conditional on \( \theta \), is a proper prior distribution that is ‘centred at \( \theta \)’, and is fully automatic so that it does not need any hyperparameter to be adjusted. Further, the method can be applied for small sample sizes (see Section 4). This is particularly important in analysing contingency tables where some tables might have some cells with either no observations or a small number of them.

It is also important to see that we are using the intrinsic prior methodology in two distinct ways. The first is to provide a means of connecting the individual studies to a meta-parameter, and in doing so create a meta-likelihood. This likelihood can be the basis of estimation or testing of the meta-parameter. For estimation, we would combine this intrinsic likelihood with a uniform prior on the meta-parameter. For testing, however, this will not work as the uniform prior is not suited for testing. So in this case we use the intrinsic methodology again, and derive an intrinsic prior for the meta-parameter. We then combine this intrinsic prior with the meta-likelihood to calculate posterior probabilities of hypotheses.

The paper is organized as follows. In Section 2 we derive the intrinsic priors for meta-analysis and construct a test for meta-independence in \( 2 \times 2 \) contingency tables. We assume that the sampling models in each of the \( m \) experiments is either a multinomial distribution with fixed table total, or a binomial experiment with one margin fixed. Section 3 shows how to estimate the meta-parameter in three real examples, using data already analysed by Cornfield [15], Gart [16], Yusuf et al. [1], Carlin [3], Efron [17] and Casella [18]. Section 4
gives evaluations of the intrinsic testing procedures, using both artificial examples and meta-analyses of contingency tables from smaller studies. Section 5 contains a discussion and there is an Appendix with some technical details.

2. THE $2 \times 2$ CASE

Suppose that $m$ experiments are developed to test the effectiveness of a new treatment with respect to an old one. For each of them a $2 \times 2$ contingency table with fixed total $n_k$ is available. If this is the only given total, then the underlying sampling distribution is multinomial. If we further assume that the totals of one margin are given, then the underlying sampling distribution is that of two binomials. In this section we derive the likelihood, intrinsic priors, and test statistics for testing independence (equality of treatments) of the meta-parameter.

2.1. Multinomial experiments

In this section we derive the likelihood and posterior distribution for estimation of the meta-parameter, and also the intrinsic prior and posterior probabilities for testing the meta-hypothesis of independence, for the case of multinomial experiments. The overall meta-experiment is

$$f(x|\theta, n) = M(x|\theta, n)$$

where $M$ represents the multinomial sampling distribution with meta-parameter $\theta = \{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\}$, and $\sum_{ij} \theta_{ij} = 1$. A particular experiment, the $k$th experiment is

$$f(x^{(k)}|p^{(k)}, n_k) = M(x^{(k)}|p^{(k)}, n_k)$$

$$\pi^D(p^{(k)}) = \text{Dirichlet}(p^{(k)}|1, 1, 1)$$

We note that if the Jeffreys prior is taken as the default prior, the results change very little. Suppressing the superscript $(k)$ unless needed, we define $x = \{x_{11}, x_{12}, x_{21}, x_{22}\}$ with $\sum_{ij} x_{ij} = n$.

The marginal $m_k(x)$ is given by

$$m_k(x) = \int M(x|p, n_k)\pi^D(p^{(k)}) \, dp = 3!C(n_k, x) \frac{\prod_{ij} x_{ij}!}{(n_k + 3)!}$$

where $C(n_k, x) = n_k!/(\prod_{ij} x_{ij}!)$, and for the comparison between the $k$th model and the meta-model, the intrinsic prior for $p^{(k)}$, conditional on $\theta$, is

$$\pi'(p^{(k)}|\theta, n_k) = E_{x|p^{(k)}} \frac{f(x|\theta, n_k)}{m_k(x)} = \sum_x \frac{f(x|\theta, n_k)}{m_k(x)} f(x|p^{(k)}, n_k)$$

(7)

where the sum is over the set of all $2 \times 2$ matrices having total $n_k$.

The relationship in (7), between the parameter $p^{(k)}$ and the meta-parameter $\theta$, is what we are looking for. We see that the intrinsic prior for $p^{(k)}$ is a weighted average of the underlying sampling models, with the weights determined by the value of the meta-parameter.
We now use the intrinsic prior (7) to calculate the likelihood of the meta-parameter $\theta$ with respect to the data $y$ (which is also the sampling distribution of $y$ conditional on the meta-parameter). Start from the likelihood of the meta-parameter $\theta$ with respect to the data $y$. An individual sample density, conditional on the meta-parameter $\theta$, can be written as

$$f(y|\theta, n_k) = \int M(y|p, n_k)\pi'(p|\theta, n_k)\,dp$$

$$= \sum_x p_M(x, y, n_k)C(n_k, x)\theta^x$$

(8)

where

$$p_M(x, y, n_k) = \frac{(n_k + 3)!2n_k!}{n!(2n_k + 3)!} \frac{C(n_k, x)C(n_k, y)}{C(2n_k, x + y)}$$

the sum is over all $x$: $\sum_{ij} x_{ij} = n_k$, and we use the notation $a^b = \prod_{ij} a_{ij}^{b_{ij}}$ for two matrices of the same dimension. From (8) we see that the sampling density conditional on the meta-parameter is a weighted average of $\theta$, with weights determined by $p_M(x, y, n_k)$. This is the probability mass function of a multinomial-Dirichlet mixture (see Section 2.3) satisfying $\sum_x p_M(x, y, n_k) = 1$.

Assuming that the data across experiments are independent, conditional on $\theta$, the likelihood of $\theta$ for the entire data set $y = \{y^{(k)}, k = 1, \ldots, m\}$ is

$$f(y|\theta, n) = \prod_k \sum_x p_M(x, y, n_k)C(n_k, x)\theta^x$$

where $n = (n_1, \ldots, n_m)$.

For computational purposes it is sometimes more convenient to work with the expression

$$f(y|\theta, n) = \sum_x P_M(x, y, n)C_M(x, n)\theta^{\sum_{x^{(k)}}}$$

where the sum is over the set of all $x = \{x^{(1)}, \ldots, x^{(m)}\}$, where $x^{(k)}$ is a $2 \times 2$ matrix with total $n_k$, and

$$P_M(x, y, n) = \prod_k p_M(x^{(k)}, y^{(k)}, n_k)$$

$$C_M(x, n) = \prod_k C(n_k, x^{(k)})$$

(10)

2.1.1. Testing. For testing the hypotheses in (5), we derive the intrinsic prior and the associated Bayes factor.

To calculate the intrinsic prior we first obtain the marginal distribution of the data under the two models $M_0$ and $M_1$ by integrating out the parameters with respect to their priors. The marginal under $H_0$ is

$$m_0(y|n) = \sum_x \frac{N!P_M(x, y, n)C_M(x, n)}{(N + 3)!C\left(N, \sum_k r_1(x^{(k)})\right)C\left(N, \sum_k c_1(x^{(k)})\right)}$$

(11)
where \( r_1(\mathbf{x}), c_1(\mathbf{x}) \) are the sums of the first row and column of the matrix \( \mathbf{x} \), respectively, and \( N = \sum_k n_k \). Likewise,

\[
m_1(\mathbf{y}|\mathbf{n}) = \sum_x \frac{N!P_M(\mathbf{x}, \mathbf{y}, \mathbf{n})C_M(\mathbf{x}, \mathbf{n})}{(N + 3)!C(N, \sum_k (\mathbf{x}^{(k)}))}
\]

(12)

The default Bayes factor is then \( m_0(\mathbf{y}|\mathbf{n})/m_1(\mathbf{y}|\mathbf{n}) \), and the intrinsic prior is the expectation of this Bayes factor with respect to the sampling density of \( M_1 \),

\[
\pi'(\theta|\mathbf{n}) = \sum_x \frac{m_0(\mathbf{y}|\mathbf{n})}{m_1(\mathbf{y}|\mathbf{n})} f(\mathbf{x}|\theta, \mathbf{n})
\]

Finally, the marginal of the data \( \mathbf{y} \) under \( M_1 \) with the intrinsic prior (13) is

\[
m'(\mathbf{y}|\mathbf{n}) = \int f(\mathbf{y}|\theta, \mathbf{n})\pi'(\theta|\mathbf{n}) d\theta
\]

\[
= \int f(\mathbf{y}|\theta, \mathbf{n})\sum_{\mathbf{y}'} \frac{m_0(\mathbf{y}'|\mathbf{n})}{m_1(\mathbf{y}'|\mathbf{n})} f(\mathbf{y}'|\theta, \mathbf{n}) d\theta
\]

\[
= \sum_{\mathbf{y}'x\mathbf{x}'} \frac{m_0(\mathbf{y}'|\mathbf{n})}{m_1(\mathbf{y}'|\mathbf{n})} P_M(\mathbf{x}, \mathbf{y}, \mathbf{n})C_M(\mathbf{x}, \mathbf{n})P_M(\mathbf{x}', \mathbf{y}', \mathbf{n})C_M(\mathbf{x}', \mathbf{n})
\]

\[
\times \int \theta^{\sum_i (x^{(i)} + x'^{(i)})} d\theta
\]

\[
= \sum_{\mathbf{y}'x\mathbf{x}'} \frac{m_0(\mathbf{y}'|\mathbf{n})}{m_1(\mathbf{y}'|\mathbf{n})} \frac{2N!P_M(\mathbf{x}, \mathbf{y}, \mathbf{n})C_M(\mathbf{x}, \mathbf{n})P_M(\mathbf{x}', \mathbf{y}', \mathbf{n})C_M(\mathbf{x}', \mathbf{n})}{2N!(2N + 3)!C(2N, \sum_k (\mathbf{x}^{(k)} + \mathbf{x}^{(k')})}
\]

(14)

The sum is over all \( \mathbf{x}, \mathbf{x}', \mathbf{y}' \) where, for example, \( \mathbf{x} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\} \) with \( \mathbf{x}^{(k)} \) a \( 2 \times 2 \) matrix with total \( n_k \).

The Bayes factor for the intrinsic prior is \( m'(\mathbf{y}|\mathbf{n})/m_0(\mathbf{y}|\mathbf{n}) \). By assuming that \( P(M_0) = P(M_1) = 1/2 \), the support that the data gives the hypothesis \( H_0 \) is now

\[
P(M_0|\mathbf{y}) = \frac{1}{1 + (m'(\mathbf{y}|\mathbf{n})/m_0(\mathbf{y}|\mathbf{n}))}
\]

(15)

2.1.2. Estimation. For estimation of the meta-parameter \( \theta \), we can start from the likelihood (9) and use a Dirichlet(\( \theta|1,1,1,1 \)) prior distribution. This would give the posterior distribution

\[
\pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta, \mathbf{n})}{\int f(\mathbf{y}|\theta, \mathbf{n}) d\theta}
\]

In practice we have found that the denominator integral is not too difficult to compute. However, for the multinomial case we have not found this distribution to be particularly useful. This is because \( \theta = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) \) is a three dimensional parameter, making it difficult to graph. Moreover, the hypothesis of independence specifies \( H_0: \theta_{11} = (\theta_{11} + \theta_{12})(\theta_{11} + \theta_{21}) \), which is difficult to assess using this posterior distribution.
2.2. Binomial experiments

For a binomial experiment, the overall meta-experiment is

\[ f(x|\theta, n) = B(x_1|n_1, \theta_1)B(x_2|n_2, \theta_2) \]

and a particular experiment, the \( k \)th experiment, is

\[ f(x^{(k)}|p^{(k)}, n^{(k)}) = B(x_1^{(k)}|n_1^{(k)}, p_1^{(k)})B(x_2^{(k)}|n_2^{(k)}, p_2^{(k)}) \]

\[ \pi(p^{(k)}) = I_{(0,1)}(p_1^{(k)})I_{(0,1)}(p_2^{(k)}) \]

For a particular experiment, the marginal distribution of \( x_i \) (suppressing the superscript if there is no ambiguity) is \( 1/(n_i + 1) \), and the intrinsic prior on \((p_1, p_2)\) is given by

\[ \pi'(p|\theta) = (n_1 + 1)(n_2 + 1)E[B(x_1|n_1, \theta_1)B(x_2|n_2, \theta_2)] \]

\[ = (n_1 + 1)(n_2 + 1)\sum_{x_1} \sum_{x_2} \left[ \prod_{i=1}^{2} B(x_i|n_i, \theta_i)B(x_i|n_i, p_i) \right] \]  

(16)

Thus, as in the multinomial case, the intrinsic prior is a weighted average of all of the possible sample values, with the weights determined by the value of the meta-parameter.

2.2.1. Meta-likelihood. As in Section 2.1, we now use the intrinsic prior (16) to calculate the likelihood of the meta-parameter \( \theta \) with respect to the data \( y \). An individual sample density, conditional on the meta-parameter \( \theta \), can be written as

\[ f(y_1, y_2|\theta, n_1, n_2) = \sum_{x_1} \sum_{x_2} \prod_{i=1}^{2} p_B(x_i, y_i, n_i)C(n_i, x_i)\theta_i^x(1 - \theta_i)^{n_i - x_i} \]

where

\[ p_B(x, y, n) = \frac{(n + 1)C(n, x)C(n, y)}{(2n + 1)C(2n, x + y)} \]  

(17)

Assuming independence of the data across the experiments, conditional on the meta-parameter, the sampling density of the entire data set \( y = \{(y_1^{(k)}, y_2^{(k)}), k = 1, \ldots, m\} \) is

\[ f(y|\theta, n) = \prod_{k} f(y_1^{(k)}, y_2^{(k)}|\theta, n_1^{(k)}, n_2^{(k)}) \]

(18)

with \( n = \{(n_1^{(k)}, n_2^{(k)}), k = 1, \ldots, m\} \). If we define

\[ P_B(x, y, n) = \prod_{k} \prod_{i=1}^{2} p_B(x_i^{(k)}, y_i^{(k)}, n_i^{(k)}) \]

\[ C_B(x, n) = \prod_{k} \prod_{i=1}^{2} C(n_i^{(k)}, x_i^{(k)}) \]

(19)
we can write

\[ f(y|\theta, n) = \sum_x \left[ P_B(x, y, n)C_B(x, n) \prod_{i=1}^2 \theta_i^{x_i^{(k)}(1 - \theta_i)^{1 - x_i^{(k)}}} \right] \]

where the sum ranges over all \( x = \{x_1^{(k)}, x_2^{(k)}\}, k = 1, \ldots, m \}. We have again put the sum outside of the product to get a more useful expression for integrating out the parameters. Similarly, under the null hypothesis \( \theta_1 = \theta_2 = \theta_0 \), the sample density is

\[ f_0(y|\theta_0, n) = \sum_x P_B(x, y, n)C_B(x, n)\theta_0^{\sum x_i^{(k)} + x_2^{(k)}} \times (1 - \theta_0)^{\sum (n_i^{(k)} + n_2^{(k)}) - (x_i^{(k)} + x_2^{(k)})} \]

### 2.2.2. Testing.

To calculate the intrinsic prior of the meta-parameter \( \theta \), and intrinsic marginal of the data, we first need the marginal densities of \( f \) and \( f_0 \) with respect to the default priors. For \( f \) we use the default product of uniform densities to obtain the marginal distribution

\[ m_1(y|n) = \sum_x \frac{P_B(x, y, n)C_B(x, n)}{\prod_{i=1}^2 (\sum_k n_i^{(k)} + 1)C(\sum_k n_i^{(k)}, \sum_k x_i^{(k)})} \tag{20} \]

and for \( f_0 \), using a uniform density on the common parameter gives

\[ m_0(y|n) = \sum_x \frac{P_B(x, y, n)C_B(x, n)}{(\sum_k n_i^{(k)} + n_2^{(k)}) + 1)C(\sum_k n_i^{(k)} + n_2^{(k)}), \sum_k (x_i^{(k)} + x_2^{(k)})}) \tag{21} \]

The intrinsic prior for the meta-parameter \( \theta \) is the expectation of the ratio of the marginal densities, that is,

\[ \pi^I(\theta|\theta_0) = E \left[ \frac{m_0(y|n)}{m_1(y|n)} \right] = \sum_y \frac{m_0(y|n)}{m_1(y|n)} f(y|\theta, n) \]

and finally the intrinsic marginal density is

\[ m^I(y|n) = \int_{\theta} f(y|\theta, n) \sum_{y'} \frac{m_0(y'|n)}{m_1(y'|n)} f(y'|\theta, n) d\theta \]

\[ = \int_{\theta} \sum_x P_B(x, y, n)C_B(x, n) \sum_{y'} \frac{m_0(y'|n)}{m_1(y'|n)} \sum_{y'} P_B(x', y', n)C_B(x', n) \times \prod_{i=1}^2 \theta_i^{x_i^{(k)}(1 - \theta_i)^{(1 - x_i^{(k)})}} \sum_{i} 2n_i^{(k)} - (x_i^{(k)} + x_2^{(k)}) \]

\[ = \sum_{xy', x' n} \frac{m_0(y'|n)}{m_1(y'|n)} \frac{P_B(x, y, n)C_B(x, n)P_B(x', y', n)C_B(x', n)}{\prod_{i=1}^2 (2N_i + 1)C(2N_i, \sum_k (x_i^{(k)} + x_2^{(k)}))} \tag{22} \]

where \( N_i = \sum_k n_i^{(k)} \).
The test of meta-independence is the test of

\[ H_0 : \theta_0 = \theta_1 = \theta_2 \] versus \[ H_1 : \theta_1 \neq \theta_2 \]

and is done by comparing the models

\[ M_0 : \{ f_\theta(y, \mathbf{n}), \pi(\theta_0) = 1_{(0,1)}(\theta_0) \} \quad \text{and} \quad M_1 : \{ f(y|\theta, \mathbf{n}), \pi'(\theta) \} \]

We calculate the Bayes factor and posterior probability as in Section 2.2.2.

2.2.3. Estimation. Estimation of the meta-parameters in the binomial case is more useful than the multinomial case. From Section 2.2.1 we can write the sample density, conditional on the meta-parameter, as

\[
f(y|\theta, \mathbf{n}) = \prod_{i=1}^{2} \left[ \prod_{k} \sum_{x(i)} p_{\theta}(x_i^{(k)}, y_i^{(k)}, n_i^{(k)}) C(n_i^{(k)}, x_i^{(k)}\theta_i^{(k)}) (1 - \theta_i)^{n_i^{(k)} - x_i^{(k)}} \right]
\]

Thus, if we put independent prior distributions on \( \theta_1 \) and \( \theta_2 \), this will result in independent posterior distributions. For the purposes of estimation, putting independent uniform priors on \( \theta_1 \) and \( \theta_2 \), resulting in the posterior distributions

\[
\pi(\theta_i|y) = \frac{f(y|\theta_i, \mathbf{n})}{\int f(y|\theta_i, \mathbf{n}) \, d\theta_i}, \quad i = 1, 2
\]

results in a reasonable posterior distribution for estimation. The denominator integral is not too difficult to compute.

2.3. Calculations

Calculation of the intrinsic marginals for testing, such as (14) or (22) is quite challenging, as the space we are summing over is huge. The only way to accomplish it is through a Monte Carlo calculation. Moreover, we need to be extremely careful in the generation of random variables. Since we expect that much of the mass is concentrated near the observed data, we use this fact to generate random variables for our Monte Carlo calculation.

2.3.1. Multinomial. Note that \( p_M(x, y, n_k) \) of (8) is the pmf of a multinomial/Dirichlet mixture. Specifically, if

\[
X|t \sim \text{Multinomial}(n, t) \quad t \sim \text{Dirichlet}(y + 1)
\]

then

\[
P(X = x) = \frac{(n + 3)(n + 2)(n + 1)}{(2n + 3)(2n + 2)(2n + 1)} \frac{C(n, x)C(n, y)}{C(2n, x + y)} = p_M(x, y, n)
\]
The function \( P_M(x, y, n) \) of (10) is the pmf of the vector \((x^{(1)}, \ldots, x^{(m)})\) conditional on \((y^{(1)}, \ldots, y^{(m)})\), where the \(x^{(k)}\) are independent with pmf \( P_M(x, y^{(k)}, n^{(k)}) \). Moreover, the density is symmetric in the sense that we can also consider it a multinomial/Dirichlet pmf of \( y \) given \( x \).

To calculate \( m_0 \) and \( m_1 \) of (11) and (12), we use an importance sampler. If \( y_{\text{obs}} \) denotes the observed data, multiply \( m_0 \) and \( m_1 \) by \( P_M(x, y_{\text{obs}}, n)/P_M(x, y_{\text{obs}}, n) \) and use the Monte Carlo sums

\[
m_0(y, n) = \frac{1}{M} \sum_j \frac{N! \text{C}_M(x_j, n)}{(N + 3)! \text{C}(N, \sum_k r_k(x_j^{(k)})) \text{C}(N, \sum_k c_1(x_j^{(k)}))} \frac{P_M(x_j, y, n)}{P_M(x_j, y_{\text{obs}}, n)}
\]

\[
m_1(y, n) = \frac{1}{M} \sum_j \frac{N! \text{C}_M(x_j, n)}{(N + 3)! \text{C}(N, \sum_k x_j^{(k)}))} \frac{P_M(x_j, y, n)}{P_M(x_j, y_{\text{obs}}, n)}
\]

where \( x_j \sim P_M(x, y_{\text{obs}}, n) \).

For the intrinsic marginal (14) we first multiply top and bottom by \( P_M(x, y_{\text{obs}}, n) \) and rearrange to get

\[
m'(y|n) = \frac{2N!}{(2N + 3)!} \sum_{y' \in \mathcal{X}} \left[ \frac{m_0(y'|n)}{m_1(y'|n)} \frac{C_M(x, n)C_M(x', n)}{C(2N, \sum_k (x^{(k)} + x'^{(k)}))} \frac{P_M(x, y, n)}{P_M(x, y_{\text{obs}}, n)} \right]
\]

\[
\times [P_M(x, y_{\text{obs}}, n)P_M(x', y', n)P_M(y', y_{\text{obs}}, n)]
\]

Conditional on \( y_{\text{obs}} \) the pmf \( P_M(x, y_{\text{obs}}, n)P_M(x', y', n)P_M(y', y_{\text{obs}}, n) \) can be considered a joint pmf of \( x, x', y' \). We therefore simulate \( x_j|y_{\text{obs}}, y_j'|y_{\text{obs}}, x_j'|y_j' \) from these distributions and use the Monte Carlo sum

\[
m'(y|n) = \frac{2N!}{(2N + 3)!} \frac{1}{M} \sum_j \left[ \frac{m_0(y'|n)}{m_1(y'|n)} \frac{C_M(x_j, n)C_M(x'_j, n)P_M(x_j, y, n)}{C(2N, \sum_k (x^{(k)} + x'^{(k)}))} \frac{P_M(x_j, y, n)}{P_M(x_j, y_{\text{obs}}, n)} \right]
\]

(23)

2.3.2. Binomial. Note that (17) is the pmf of a beta-binomial random variable. Specifically, if

\[X|t \sim \text{Binomial}(n, t)\]

\[t \sim \text{Beta}(y + 1, n - y + 1)\]

then

\[P(X = x) = \frac{(n + 1) C(n, x) C(n, y)}{(2n + 1) C(2n, x + y)} = p_B(x, y, n)\]
Furthermore, the function $P_B(x, y, n)$ of (19) is the pmf of the vector $(X_i, X_j) = ((X_i^{(1)}, \ldots, X_i^{(m)}), (X_j^{(1)}, \ldots, X_j^{(m)}))$ conditional on $(Y_1, Y_2)$, where the $X_i^{(k)}$ are independent with pmf $p_B(x, y_i^{(k)}, n_i^{(k)})$. Note also that (19) is symmetric in that we can consider it a beta-binomial density of $Y$ conditional on $X$.

For the calculations here we again use an importance sampler; for $m_1$ and $m_0$ of (20) and (21) generate $x_j \sim P_B(x, y_{\text{obs}}, n)$ and use the Monte Carlo sums

$$m_1(y|n) = \frac{1}{M} \sum_j \frac{C_B(x_j, n)}{\prod_{i=1}^2 (\sum_k n_i^{(k)} + 1) C(\sum_k n_i^{(k)}, \sum_k x_i^{(k)}) P_B(x_j, y, n)}$$

and

$$m_0(y|n) = \frac{1}{M} \sum_j \frac{C_B(x_j, n)}{(\sum_k n_1^{(k)} + n_2^{(k)}) + 1) C(\sum_k n_i^{(k)}, \sum_k x_i^{(k)}) P_B(x_j, y, n)}$$

Similarly, for the intrinsic marginal (22), we multiply top and bottom by $P_B(y', y_{\text{obs}}, n)P_B(x, y_{\text{obs}}, n)$ and simulate $x_j|y_{\text{obs}}, y'_{\text{obs}}, x'_j|y'$ for the Monte Carlo sum

$$m'(y|n) = \frac{1}{M} \sum_j \frac{m_0(y'_j|n)}{m_1(y'_j|n)} \frac{C_B(x_j, n) C_B(x'_j, n)}{\prod_{i=1}^2 (2N_i + 1) C(2N_i, \sum_k x_i^{(k')} + x'_i^{(k)}) P_B(x'_j, y_{\text{obs}}, n)}$$

$$\times \frac{P_B(x'_j, y, n)}{P_B(x'_j, y_{\text{obs}}, n)}$$

3. ESTIMATION OF THE META-PARAMETER

In this section we present three meta-analyses on known data sets. The first two have had a number of analyses performed on them, and the latter less so. The first two data sets, the smoker data and the beta-blocker data, yield very different evidence about the meta-parameter. Here we use the estimation approach, and estimate the posterior distribution of each success probability in the two-binomials case.

The first data set consists of 14 case-control studies on lung cancer patients. In each study the number of non-smokers among lung cancer patients (cases) and the number of non-smokers among non-cancer patients (controls) was recorded. The data have a long history, and can be found in Reference [19, Section 2.7.3] or Reference [3]. The second data set, the beta-blocker data, (which also can be found in Reference [3]) consists of 22 clinical trials of beta-blockers for reducing the mortality after myocardial infarction, where the data are the number of deaths and totals for the population under treatment and the number of deaths and total for the control population. The third example, the ulcer data, can be found in Reference [17] or Reference [18], and consists of 39 studies of a surgical treatment (new versus old) for ulcers, with the success being a measure of non-recurrence of bleeding.

As a first look at what is going on, we offer Figure 1, which shows individual $p$-values and intrinsic posterior probabilities for these three studies. (These are posterior probabilities of the individual studies, calculated according to Appendix A.) As can be seen, the studies
in the smoker data set are mostly significant, leading to an overall conclusion that the null hypothesis should be rejected. In the beta-blocker data the opposite is true, in that most of the studies are non-significant, and the overall conclusion seems to be to accept the null hypothesis (or, as has been done, to conclude that there is a small treatment effect). Lastly, the ulcer data have a mix of significant and non-significant studies, and no clear conclusion. We note that in both the beta-blocker and ulcer data, the posterior probabilities give slightly more evidence than the $p$-value for the presence of a treatment effect.

Thus, according to the individual measurements, the conclusion from the ulcer data seems least clear. When we do a meta-analysis on these data sets, in all cases the conclusions become quite clear, and, somewhat surprisingly, even though the individual studies do not seem to give clear evidence for the ulcer data, the meta-analysis is very conclusive.

Figure 2 shows all of the posterior distributions calculated according to the methods of Section 2.2.3, and Table I gives summary statistics of the distributions. It is interesting that in every case the conclusion is clear. The smoker data and, to a lesser extent, the ulcer data, result in posterior densities that do not overlap, showing a clear treatment effect. This is also reflected in the summary statistics, and the non-overlapping HPD regions. For the beta-blocker data the conclusions from the graphs of the posterior distributions are less clear, but are made a bit clearer by the summary statistics. There we see the non-overlap (be it slight!) of the 95
Figure 2. Posterior densities for treatments or cases (solid lines) and controls (dotted lines) for the smoker data (left panel), beta-blocker data (middle panel), and the ulcer data (right panel).

### Table I. Summary statistics for the posterior distributions from the three meta-analyses.

<table>
<thead>
<tr>
<th></th>
<th>Mode</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>95 per cent HPD region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker Cases</td>
<td>0.833</td>
<td>0.833</td>
<td>0.007</td>
<td>(0.819, 0.848)</td>
</tr>
<tr>
<td>Control</td>
<td>0.976</td>
<td>0.976</td>
<td>0.004</td>
<td>(0.969, 0.983)</td>
</tr>
<tr>
<td>Beta-blocker</td>
<td>0.076</td>
<td>0.076</td>
<td>0.005</td>
<td>(0.068, 0.085)</td>
</tr>
<tr>
<td>Control</td>
<td>0.096</td>
<td>0.096</td>
<td>0.005</td>
<td>(0.086, 0.106)</td>
</tr>
<tr>
<td>Ulcer Trt.</td>
<td>0.833</td>
<td>0.831</td>
<td>0.023</td>
<td>(0.786, 0.880)</td>
</tr>
<tr>
<td>Control</td>
<td>0.642</td>
<td>0.641</td>
<td>0.029</td>
<td>(0.583, 0.670)</td>
</tr>
</tbody>
</table>

per cent HPD regions, leading us to conclude that there is a treatment effect, although most would probably call this a ‘slight’ treatment effect.

4. TESTING THE META-PARAMETER

In the previous section we illustrated the use of the meta-likelihood for estimation of the meta-parameter, and saw the power of the method. We next turn to the problem of testing the meta-parameter.
There are a number of reasons why one might prefer a testing approach to an estimation approach. To us, meta-analysis is, fundamentally, a testing problem, so we are most comfortable with that approach. However, the testing formulation brings along a high computational overhead, much more so than estimation. Thus, estimation is an easier route if it is applicable. However, in certain cases testing is necessary. In particular, if the data are multinomial (which is a case that can be made for the ulcer data), the estimation approach will not work. Also, if the estimation is not clearcut, a testing conclusion may be preferred.

This section examines the performance of the intrinsic posterior probabilities in a number of different ways. First we look at some artificial examples to better understand the intrinsic posterior probabilities, then we analyse some smaller meta-analyses, and see that the intrinsic method can be useful even if there are only a few studies with small counts.

4.1. Evaluations

In this section we evaluate the performance of the intrinsic posterior probability under a number of different scenarios, considering artificial examples in which the underlying structure is clear. We present these examples to demonstrate that the meta-density and posterior probabilities function as expected in problems in which the answers are clear.

4.1.1. The sample density conditional on the meta-parameter

The sample density (8), the density of the sample conditional on the meta-parameter, might appear unusual at first sight. The density is a mixture over all possible tables, with the data contributing to the weight that a particular table has. The exponent of the parameter does not contain the data, as would be the case with a typical multinomial likelihood.

To verify that, indeed, this density behaves as we would want it to, we present a small example. For a $2 \times 2$ table with $n = 8$, it is possible to calculate the density exactly (there are 165 possible tables). Figure 3 shows a plot of the density for parameter configurations $\theta = \{0.25, 0.25, 0.25, 0.25\}$ and $\theta = \{0.1, 0.4, 0.4, 0.1\}$. The density behaves exactly as desired. In the first case the table $\{2,2,2,2\}$ gets the highest weight, and in the second case the table $\{0,4,4,0\}$ receives the highest weight. Moreover, the density gives less weight to tables that are less likely under the given parameter configurations. For example, if $\theta = \{0.25, 0.25, 0.25, 0.25\}$, the table $\{8,0,0,0\}$ is very unlikely, and receives the smallest weight.

4.1.2. Some artificial tables.

To demonstrate that the intrinsic posterior probability is functioning as it should, and providing the inference that we desire, we provide a small example using tables for which the conclusion is ‘obvious’. For example, for a meta-analysis with three experiments, each of which results in a $2 \times 2$ contingency table with each cell having the number 3, it is fairly clear that the null hypothesis of independence is supported. Table II shows the results of four such sets of tables. As we read down Table II, the sets of artificial tables are constructed to provide increasing evidence against the null hypothesis. This is confirmed by the decreasing trend in both posterior probabilities.

Figure 4 shows the prior distribution for the experiment described in Table II. The prior distribution only depends on the number of trials done in each experiment (which can be seen from (7)), and is symmetric around the line $\theta_1 = \theta_2$. The four posterior distributions shown in Figure 5 depict increasing skewness, and increasing support for the alternative as the observed tables become more extreme.
Figure 3. Density (8) for the $2 \times 2$ table with $n = 8$. The upper panel has $\theta = \{0.25,0.25,0.25,0.25\}$ and the lower panel has $\theta = \{0.1,0.4,0.4,0.1\}$. The tables are ordered according to their probability under $\theta = \{0.25,0.25,0.25,0.25\}$. Note that in the upper panel only distinct tables will have different probabilities.

Table II. Four sets of artificial tables, with the binomial and multinomial meta posterior probability.

<table>
<thead>
<tr>
<th></th>
<th>Tables</th>
<th>Binomial meta-post. prob</th>
<th>Multinomial meta-post. prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\begin{bmatrix} 3 &amp; 3 \ 3 &amp; 3 \end{bmatrix}$</td>
<td>0.564</td>
<td>0.627</td>
</tr>
<tr>
<td>(b)</td>
<td>$\begin{bmatrix} 3 &amp; 3 \ 3 &amp; 3 \end{bmatrix}$</td>
<td>0.442</td>
<td>0.558</td>
</tr>
<tr>
<td>(c)</td>
<td>$\begin{bmatrix} 3 &amp; 3 \ 3 &amp; 3 \end{bmatrix}$</td>
<td>0.162</td>
<td>0.106</td>
</tr>
<tr>
<td>(d)</td>
<td>$\begin{bmatrix} 6 &amp; 0 \ 0 &amp; 6 \end{bmatrix}$</td>
<td>0.009</td>
<td>0.004</td>
</tr>
</tbody>
</table>

4.2. Meta-analyses

Here we look at two small data sets for meta-analysis. We chose small data sets for two reasons. First, we wanted to see the power of the intrinsic posterior probabilities on small
tables. Second, as we noted previously, the computational burden is high, and calculations on small tables are more feasible. Moreover, with large counts the estimation methods of Section 3 are typically quite conclusive.

First we look at a subset of the ulcer data (Table III), where we chose tables with small counts. Also, in this analysis, we are able to treat the $2 \times 2$ tables using the multinomial distribution, an assumption that more closely fits the experiment. Second, we look at a subset of the PEP data (Table IV) analysed by George et al. [20]. Patients undergoing endoscopy are given one of two contrast media, a high osmolality contrast media (HOCM) or a low osmolality contrast media (LOCM). The question of interest is whether the type of contrast media is associated with a toxic renal event called PEP (Post-ERCP Pancreatitis), which is an unfavourable outcome. Although use of LOCM has been postulated as safer, the evidence has not been conclusive. Moreover, the cost of LOCM is approximately 20–40 times that of HOCM, although both media provide the same image quality. George et al. [20] undertook a meta-analysis of 14 published studies that compared HOCM and LOCM.

**Ulcer data:** From Table III we see that both the individual $p$-values and the individual posterior probabilities are inconclusive, as some are significant and some are not. However, the posterior probability of the meta-hypothesis, calculated using (23), is decisive. When the seven studies are included, the posterior probability of $H_0$ is 0.101, giving decisive evidence for a treatment effect (the same conclusion reached when all 39 studies were used).

**PEP data:** From Table IV it is also the case that neither the individual $p$-values or posterior probabilities give a clear conclusion. However, again the meta-posterior probability (calculated using (A1)) is decisive, giving a posterior probability of $H_0$ of 0.675 when five studies are included in the analysis. Thus, there is no evidence of a treatment effect, and HOCM can be used in the place of the more costly LOCM, which is the same conclusion reached by George et al. [20] using 14 studies.

Figure 4. Meta prior distribution for the experiments in Table II. Note that the prior only depends on the number of binomial trials done in each experiment.
Figure 5. Posterior distributions for the data in Table II. The upper left is the posterior for the data (a); the upper right for (b); the lower left for (c); and the lower right for (d).

Table III. Ulcer data (subset). Data are (Yes, No) for the occurrences of an ‘event’–non-occurrence of bleeding. The meta-posterior probabilities are cumulative, each row includes all previous rows.

<table>
<thead>
<tr>
<th>(New, old)</th>
<th>p-value</th>
<th>Indiv. post. prob</th>
<th>Meta-post. prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4,7)</td>
<td>0.278</td>
<td>0.522</td>
<td>0.444</td>
</tr>
<tr>
<td>(6,6)</td>
<td>0.194</td>
<td>0.478</td>
<td>0.474</td>
</tr>
<tr>
<td>(4,3)</td>
<td>0.125</td>
<td>0.399</td>
<td>0.488</td>
</tr>
<tr>
<td>(2,5)</td>
<td>0.049</td>
<td>0.301</td>
<td>0.325</td>
</tr>
<tr>
<td>(6,6)</td>
<td>0.016</td>
<td>0.165</td>
<td>0.315</td>
</tr>
<tr>
<td>(2,8)</td>
<td>0.007</td>
<td>0.124</td>
<td>0.248</td>
</tr>
<tr>
<td>(0,6)</td>
<td>0.0005</td>
<td>0.019</td>
<td>0.101</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The model for a meta-analysis is a hierarchical model in which the interest is the parameters at the deepest level of the hierarchy. These are the parameters that are common to all studies, the
meta-parameters. By using the intrinsic prior methodology we have constructed a hierarchical model for meta-analysis of contingency tables. The performance, on both real and artificial examples, has shown that the method is a reasonable one.

5.1. The intrinsic hierarchy

For models such as hierarchical linear models, the structure of the hierarchy is typically clear. At each level the parameters are the means of the previous level. However, for contingency table data the method of structuring the hierarchy is less clear. One of the more common methods is to use a linear model type structure on a transform of the contingency table parameters, for example on the logit.

If a transformation, such as a logit, together with a normal (or closely related) hierarchy is not used, but instead one chooses to model the data as collected (as multinomial or binomial), then there is no simple model to connect the parameters. For example, with a binomial sampling distribution and a beta prior on \( p^k \), it is virtually impossible to connect \( p^k \) to \( \theta \). However, we can use the intrinsic prior methodology to connect the levels of the hierarchy.

The intrinsic prior construction can be simply interpreted as a multiple testing problem where the null (the meta-model) is nested in each of the experiments. Under this assumption, the method outlined in Section 1 will allow us to construct good prior distributions first for \( p^k \) conditional on \( \theta \), and then unconditionally for \( \theta \).

The resulting intrinsic priors have a very interesting structure, which may indicate why we might expect good performance. From (A1) and (A2) it can be seen that the intrinsic marginal is the expected value of \( m_0 \) under the predictive distribution. Thus, we are using a prior that reflects the null hypothesis, but also takes into account the variability in the problem.

This is a very natural method, and actually can be made quite general. The calculation of the intrinsic prior is straightforward and does not depend on the observed data, but only on a theoretical training sample and the sampling distribution. Indeed, this approach has somewhat of a flavour of a missing data problem, where we consider the parameters at one level of a hierarchy to be missing data from the deeper level.

5.2. Data analyses

The data sets analysed here highlight very different aspects of a meta-analysis. First, the conclusions from the smoker data are quite clear. The effect is large, and all analyses have led to the same conclusion.

The beta-blocker data are quite a different situation, and different conclusions have been reached. The analysis of Yusef et al. [1], using the method of Peto, concludes that there is
strong evidence of a 20 per cent reduction in mortality due to beta-blockers. Such a strong conclusion is not shared by others. In particular, Carlin [3], using a hierarchical logistic Bayes model, concludes that there is only a modest reduction in the mortality rate due to the beta-blockers. This difference in conclusions is to be somewhat expected, as Yusef et al. use a fixed effects model, while the analysis of Carlin is closer to a random effects model [2].

The intrinsic meta-analysis, being based on a hierarchical model, becomes a random effects analysis, and our conclusions are more in line with those of Carlin. The mixing that takes place when the meta-likelihood is calculated, such as in (9), is the same type of calculation that is done to obtain the likelihood in a random effects model. This is the correct way in which to do a meta-analysis; if the variability due to studies is not accounted for, as is the case in the Yusef et al. fixed effect analysis, the variability of the estimate is underestimated and Type I errors become more likely.

The analysis of the ulcer data in Section 3 was quite straightforward, and perhaps was the best illustration of the worth of a meta-analysis. The individual p-values and posterior probabilities were inconclusive, but the meta-calculation was very clear.

Lastly, the testing-based analysis of the smaller data sets in Section 4 showed the power of the intrinsic meta-posterior probability. There, we were able to reach the same conclusions as an analyses that used many more studies with more observation. This, perhaps, was the most striking evidence in favour of the intrinsic meta-analysis.

APPENDIX A: INDIVIDUAL CALCULATIONS

In this section we give details for the calculation of the individual posterior probabilities used in Section 3.

A.1. Multinomial posterior probabilities

For an individual multinomial test we compare

\[ H_0 : f(y | \theta) = C(n, y) \theta_1^{y_{11} + y_{12}} (1 - \theta_1)^{y_{21} + y_{22}} \theta_2^{y_{21} + y_{22}} (1 - \theta_2)^{y_{12} + y_{22}} \]

\[ \pi_0 = \text{beta}(1, 1) \times \text{beta}(1, 1) \]

with

\[ H_1 : f_1(y | \theta) = C(n, y) \theta^5, \pi_1 = \text{Dirichlet}(1, 1, 1, 1) \]

It is straightforward to calculate the marginal distributions under each hypothesis. We have

\[ m_0(y) = \frac{C(n, y)}{(n + 1)^2 C(n, y_{11} + y_{12}) C(n, y_{11} + y_{21})} \]

\[ m_1(y) = \frac{n!}{(n + 3)!} \]
The intrinsic marginal is then

\[
m^I(y) = \int_{\Theta} \left[ \sum_x \frac{m_0(x)}{m_1(x)} f_1(x|\theta) \right] f_1(y|\theta) \pi_1(\theta) \, d\theta
\]

\[
= \sum_x m_0(x) \left[ \frac{(n+3)!2n!}{n!(2n+3)!} C(n,x)C(n,y) \right] \left[ \frac{n!C(n,x+y)}{C(2n,x+y)} \right]
\]

(A1)

where, as in Section 2.3, the quantity in square brackets is the pmf of a Multinomial/Dirichlet mixture. Calculation of \(m^I\) is done by simulating from this distribution.

We also note that the multinomial/Dirichlet mixture is, in fact, the predictive distribution under the model in \(H_1\). Therefore, we see that the intrinsic prior is the expected value of \(m_0\) under the predictive distribution.

A.2. Binomial posterior probabilities

Calculation of the individual binomial posterior probabilities is similar to the multinomial case; we only sketch the details here.

For an individual binomial test we compare

\[
H_0 : f_0(y|\theta_0) = C(n_1, y_1)\theta_0^{y_1} (1 - \theta_0)^{n_1-y_1} C(n_2, y_2)\theta_0^{y_2} (1 - \theta_0)^{n_2-y_2}
\]

\[
\pi_0 = \text{beta}(1,1)
\]

with

\[
H_1 : f_1(y|\theta) = C(n_1, y_1)\theta_1^{y_1} (1 - \theta_1)^{n_1-y_1} C(n_2, y_2)\theta_2^{y_2} (1 - \theta_2)^{n_2-y_2}
\]

\[
\pi_0 = \text{beta}(1,1) \times \text{beta}(1,1)
\]

The marginal distributions are

\[
m_0(y) = \frac{1}{n_1 + n_2 + 1} \frac{C(n_1, y_1)C(n_2, y_2)}{C(n_1 + n_2, y_1 + y_2)}
\]

\[
m_1(y) = \frac{1}{(n_1 + 1)(n_2 + 1)}
\]

and the intrinsic marginal is

\[
m^I(y) = \sum_x m_0(x) \left[ \prod_{i=1}^2 \left( \frac{n_i + 1}{2n_i + 1} \right) \frac{C(n_i,x_i)C(n_i,y_i)}{C(2n_i,x_i+y_i)} \right]
\]

(A2)

where, as in Section 2.3, the quantity in square brackets is the pmf of a Binomial/beta mixture. Calculation of \(m^I\) is done by simulating from this distribution, and as in the previous section, the intrinsic prior is the expected value of \(m_0\) under the predictive distribution.
ACKNOWLEDGEMENTS

The work of first author was supported by National Science Foundation Grant DMS-9971586, and work of the second author was supported by Ministerio de Ciencia y Tecnologia, Grant BEC2001-2982.

REFERENCES