Bayesian Meta-Analysis
A Practical Introduction

Robert Grant and Gian Luca Di Tanna


CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Bayesian Meta-Analysis

"…this book is extremely timely…not just a technical exposition, but provides practical guidance about using different software platforms, as well as valuable advice about extracting summary statistics, eliciting prior information, communicating results, visualisation, and many other issues…reflects years of thoughtful experience, and should be of huge value to anyone faced with pooling studies into a coherent whole."

~From the Foreword by Professor Sir David Spiegelhalter

Meta-analysis is the statistical combination of previously conducted studies, often from summary statistics but sometimes with individual participant data. It is widespread in life sciences and is gaining popularity in economics and beyond. In many real-life meta-analyses, challenges in the source information, such as unreported statistics or biases, can be incorporated using Bayesian methods. *Bayesian Meta-Analysis: A Practical Introduction* provides an approachable introduction for researchers who are new to Bayes, meta-analysis, or both. There is an emphasis on hands-on learning using a variety of software packages.

Key Features

- Introductory chapters assume no prior experience or mathematical training and are aimed at non-statistical researchers.
- Examples of basic meta-analyses in seven different software alternatives: BUGS, JAGS, Stan, bayesmeta, brms, Stata, and JASP.
- Practical advice on extracting information from studies, eliciting expert opinions, managing project decisions, and writing up findings.
- Discussion of specific problems, including publication bias, unreported statistics, and a mixture of study designs, with code examples.
- Accompanying online blog and forum, with all code and data from the book, plus more translations to different software.

This book aims to bridge the gap between the researcher who wants to carry out tailored meta-analysis and the techniques they need, which have previously been available only in mathematically or computationally demanding publications.

Contents

Foreword

Scientific studies should never be seen in isolation—they add to existing knowledge, enabling us to learn and progress. This is the essence of Bayesian thinking; on receipt of new evidence, basic probability theory is used to update our beliefs in a coherent way. Meta-analysis—where the evidence from multiple studies is combined—is, therefore, a natural area for Bayesian methods, and so this book is extremely timely.

But of course things are rarely simple. Numerous complexities arise in trying to have a consistent approach to combining multiple sources of evidence with our previous understanding. Studies can vary for unknown reasons, use different designs, have missing information, or be subject to publication bias. Fortunately, a Bayesian approach allows extraordinary flexibility to non-standard features, which can otherwise be very difficult to handle.

The authors of this book have faced up to this complexity with relish, and systematically worked through a wide range of challenges encountered by those wanting to combine evidence in the real world. And this is not just a technical exposition, but provides practical guidance about using different software platforms as well as valuable advice about extracting summary statistics, eliciting prior information and communicating results, visualisation, and many other issues they and others have had to deal with. There is a wealth of useful guidance, whether adopting a Bayesian approach or not.

Although Bayesian methods are a natural fit for meta-analysis, their adoption has been surprisingly slow. Many of the challenges this book addresses were highlighted in seminal works from the 1990s and discussed extensively in the 2000s. However, currently only a small fraction of meta-analyses employ Bayesian approaches to address these issues. In contrast, Bayesian software has advanced remarkably, offering several powerful tools that this book explores. There has never been a better time to embrace Bayesian methods in meta-analysis.

This book reflects years of thoughtful experience, and should be of huge value to anyone faced with pooling studies into a coherent whole.

Professor Sir David Spiegelhalter, FRS OBE

Preface

Meta-analysis is a statistical tool to combine the results reported by a collection of similar studies. The aim is to bring clarity to decision-makers (including the public), instead of expecting them to find and reconcile multiple studies. Peter Morgan, then scientific editor of the *Canadian Medical Association*, put it like this in 1986:

> The medical literature can be compared to a jungle. It is fast growing, full of dead wood, sprinkled with hidden treasure and infested with spiders and snakes.
> [...]
> Review articles will become increasingly popular as the size of the jungle of medical literature doubles every 10 years. The number of review journals and books continues to increase as more authors learn how to use the computer to search the literature. Writing review articles will be more competitive, but it also will be more rewarding [...] [171]

This book aims to help you continue that trend toward making sense of the literature jungle, and to use computer power effectively for this.

Unfortunately, the studies that have been done on a particular topic (which we call the *evidence base*) are not always very similar, and not always very well done and/or reported. Little problems crop up that prevent us from comparing like with like, because they cannot be accommodated in the usual meta-analysis methods.

The researcher then has a difficult choice of whether to make some bold assumption to simplify the problem, or to discard potentially useful studies, just to keep the meta-analysis show on the road. This book introduces a third option: to use Bayesian methods instead, which can include many difficult features in the evidence base in a more tailored statistical model, and allow useful ways of presenting results.

We have written this book for the majority of people doing meta-analysis today: researchers who understand systematic reviews, and perhaps simple descriptive and inferential statistics, and who now need to combine statistics from other people's studies. The book is subtitled "A Practical Introduction", a task we take very seriously. We provide not just formulas, but code and examples, to get you started, and advice on how meta-analysis can get tricky when the theory meets a real evidence base. We are opinionated, but we tell you what the opposing views are, too.

We expect that most readers will be new to Bayesian methods, and so we present multiple software options, especially in Part 1 of the book. The purpose of this is to let the reader compare them and decide which they would prefer to use.

Our examples are mostly drawn from biomedical research, as that is our background and the subject of most meta-analysis today, but we also reflect on research and policy in economics and education, where meta-analysis is growing rapidly in popularity. We have tried to keep all case studies simple so that experience of the substantive topic is not required. As this book is both practical and an introduction, we do not devote any space to history and very little to the philosophy of probability; these are interesting subjects, but belong elsewhere.

We mostly consider meta-analysis of studies that compare "arms" or groups, principally randomised experimental studies, but there is some consideration given to studies that cannot count on random allocation. Meta-analysis of diagnostic or prognostic models is an important topic, but requires more space than we can accommodate here. We also do not attempt to include borrowing of external information in clinical trials, or adaptive trial designs. Research designs other than trials that lie outside our scope include pharmacokinetic and pharmaco-dynamic research, Bayesian belief networks, hidden Markov models, and differential equation methods such as in infectious disease modelling.

Part 1 of this book contains a primer on the foundations of statistical inference (Chapter 1). Readers already familiar with statistics and meta-analysis should skim, but not skip, this chapter, because it introduces our terminology and notation, as well as a way of conceptualising analyses not found in introductory statistics textbooks. Bayesian statistics and software is introduced in Chapter 2. We then introduce meta-analysis, including Bayesian equivalents to common effect models, in Chapter 3 (often called "fixed effect"), and random effects models in Chapter 4.

Part 2 considers the inputs that are essential for Bayesian meta-analysis—extracting statistics from published studies, and obtaining prior distributions, including the opinions of experts—as well as how to present outputs. Part 3 explores specific problems and how they can be modelled. Each Part 3 chapter is as short and specific as possible so that once readers have covered the basics they can use this as a practical reference guide.

Throughout, we aim to start each chapter with a motivating problem, consider simple models of the data-generating process, and show the readers enough code that they can explore the problems and get a deeper intuition. At this stage, we keep the terminology basic and the language intuitive and informal. Once the problem is understood, we propose more complicated models to deal with it. Formal definitions, if needed at all, come at the end when the reader has reached the deepest understanding*.

We introduce seven software options in Part 1: BUGS, JAGS, Stan, Stata, bayesmeta, brms, and JASP. BUGS is still the most widely used software for Bayesian meta-analyses, so we mostly present BUGS code for models in Parts 2 and 3, but the accompanying website at **`https://bayesian-ma.net`** provides translations where possible; we use base (not "tidyverse") R as a *lingua franca*. The website includes all data and code from the book. The computer symbol in the margin† is there to encourage readers to play with data and models to gain deeper understanding.

We emphasise the responsibility on the meta-analyst (and any Bayesian analyst) to make modelling choices and be prepared to explain and justify them. Readers may find themselves on a miniature version of the famous Dunning-Kruger curve of confidence when learning and practising statistics (see Figure 0.1). Arriving at the right-hand side of the curve requires mathematical mastery‡, a deep understanding of why you are doing meta-analysis, and perhaps most importantly, a critical and curious mindset.

Although our emphasis is on mastery rather than a statistical cookbook, or theorems and proofs, we have maintained the spirit of mathematical rigour in that a term or symbol only ever means one thing. We remember how small ambiguities in notation and coding can leave the novice confused and dispirited. This means that our mathematical notation can seem complicated at first, but pays dividends later.

*Michael Greenacre's books on correspondence analysis were the inspiration for short chapters, and Lara Alcock's on learning mathematics informed the structure within each chapter.

†Icon produced by Linux GNOME Project, CC-BY-SA-3.0.

‡This is a concept widely discussed in primary and secondary school mathematics, but we feel it also applies to adult professionals who are moving into unfamiliar mathematical concepts and who will need to continually adapt what they learn to new challenges.

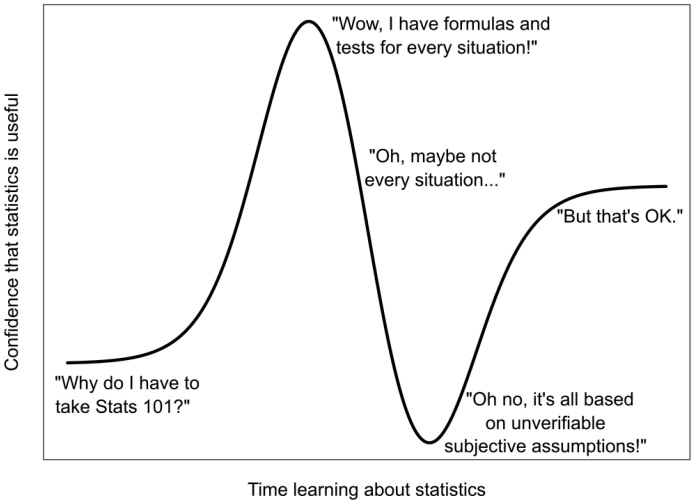

FIGURE 0.1
Our take on the Dunning-Kruger curve, based on our experience as students, then teachers of statistics. Learning Bayesian meta-analysis involves another roller-coaster ride.

Finally, we must warn readers that most people new to Bayesian methods encounter some frustrations in the early days. To have flexibility that allows tailored models, the software has to use simulation algorithms. Sometimes they will struggle, and it can be hard work to track down errors in your code or to set it up in the best way to get it running smoothly. We suspect that most new Bayesians at some point question whether it is worthwhile. We think it is. As John Tukey wrote more than 60 years ago:

> What of the future? The future of data analysis can involve great progress, the overcoming of real difficulties, and the provision of a great service to all fields of science and technology. Will it? That remains to us, to our willingness to take up the rocky road of real problems in preference to the smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments. Who is for the challenge? [260]

We hope you will find Bayesian meta-analysis to be as useful as we have, and join us along the rocky road.

Robert & Gian Luca, Hampshire & Ticino, 2024

Acknowledgements

The development of this book has been profoundly shaped by our educational experiences, both as instructors and as perpetual learners.

We extend our sincere appreciation to the students who have participated in our various courses and seminars over the years. Their engagement, inquiries, and perspectives have significantly influenced our approach in this book. Thank you, for what we learnt from you.

Anonymous reviewers gave their time and expertise generously. Critical friends and experts helped refine the book, and suggested ideas and examples; thank you: Jay Brophy, Sylwia Bujkiewicz, Paul-Christian Bürkner, Meghan Cain, Giovanni Cerulli, Sofia Dias, Andrew Gelman, Nicholas Latimer, Rachael Meager, Emilia Riggi, Eric-Jan Wagenmakers, Ian White, and Vicki Yorke-Edwards.

We are deeply indebted to colleagues in research whose guidance and support have been invaluable in creating the time and opportunity to grow and refine our ideas. In particular, we want to thank: Luca Crivelli, Vari Drennan, Gary Globe, Mike Hurley, Derek Lowe, Jaime Miranda, Tom Quinn, Eva Segelov, Jill Stoddart, Armando Teixeira-Pinto, Giorgio Treglia, Maria Grazia Valsecchi, and Annarita Vestri.

We would also like to thank Kingston University, in particular Ann Ooms and James Denholm-Price, for hosting us during a productive writing week in June 2024.

Our profound appreciation extends to Joseph Alvin Ramos Santos, who transcends the role of researcher to that of a valued friend.

4

Random Effects Meta-Analysis and Heterogeneity

Learning objectives

After reading this chapter, you will be able to:

1. justify a choice of common effect or random effects models
2. interpret random effects meta-analysis as a form of hierarchical (multilevel) regression model
3. choose, and justify, prior distributions on the unknown quantities in the model
4. use Bayesian software to conduct a random effects meta-analysis
5. write Bayesian meta-analysis models that are close matches to the DerSimonian-Laird and Sidik-Jonkman methods
6. critically consider the shape of heterogeneity in the evidence base, comparing different distributions, and justify your choice in a model
7. critically investigate heterogeneity and justify your choice of modelling subgroups in the evidence base, or meta-regression

Each study in a meta-analysis will have collected data from a "population". The population might be people with the relevant health condition in the local hospital, or on some national register. It might not be people but instead care homes or schools, or even individual blood samples.

We saw in Chapter 3 how studies are performed in different parts of the world, at different times, and they are likely to have slightly different inclusion and exclusion criteria. This means that they are each drawing data from a somewhat different population, even before any intervention or follow-up. Then, there can be slight differences in how interventions are implemented and how outcomes are defined and measured, which can alter the intervention effect.

It is reasonable to expect such studies to arrive at somewhat different results. This inter-study variation, which is called *heterogeneity*, applies in addition to the sampling distribution.

In the early days of meta-analysis, this was a source of much concern: that we should not inform clinical decisions by averaging studies when we are not comparing like with like. "Comparing apples and oranges", people often say. The controversial psychologist Hans Eysenck was an early sceptic of meta-analysis, and called such combinations of studies not meta-analysis but "mega-silliness"[70].

Although we think he was wrong to condemn the majority of meta-analyses, based on early experiences in the questionably reproducible field of personality and intelligence scales,

DOI: 10.1201/9781003375821-4

we do think that he highlighted several problems that continue to trouble meta-analysts [71]. Bayesian approaches offer a solution to most of these.

In everyday practice, we have to consider the size of the heterogeneity, compared to the sampling distributions. We cannot simply refuse to meta-analyse any evidence base that exhibits even a small degree of heterogeneity, or we will be failing to help our audience. On the other hand, we cannot put a mismatched collection of statistics into one melting pot and expect it to be helpful.

We also have to think about what might be causing studies to arrive at heterogeneous results. If we can understand that, we should describe it, and if we can go further and quantify it, then we should at least consider whether we can adjust for it. This is the focus of later chapters, but for now, we will simply set out how we can assess the size and shape of heterogeneity and thus arrive at an estimate of the underlying effect and its uncertainty.

There are two approaches in widespread use to this question of how comparable the studies are, and a third that is much rarer. We will now show how they are defined and what the impact is on the meta-analysis.

1. *Common effect* meta-analysis works on the basis that all studies have drawn participants at random from the same shared population. The objective is to estimate the intervention effect (or other relevant statistic) in that population. This was covered in detail in Chapter 3.
2. *Random effects* meta-analysis assumes that the studies are themselves drawn from a population of possible studies, and then their participants' data are in turn drawn from a study-specific population. The random effects meta-analysis allows for heterogeneity, including it as a prior distribution of the study-specific intervention effects.
3. *Fixed effects* meta-analysis is rarer, and allows for differences between studies, but estimates a intervention effect in each study's population. The overall intervention effect from the meta-analysis is then a weighted average of these, without any claim to represent a population of potential studies.

Unfortunately, many researchers that we have met have some misunderstandings about what these different approaches entail. Often, it is said—incorrectly—that a common effect meta-analysis is appropriate when there is little or no heterogeneity, and a random effects should be used instead when heterogeneity is present. Worse yet, you might encounter the suggestion that the type of meta-analysis is chosen on the basis of some measure of heterogeneity. A clear description of these options, along with debunking other myths, is given by Borenstein [23].

You might feel that the notion of study designs and populations being drawn at random from a distribution of possible studies is far-fetched. We sympathise with this intuition, but experience has shown us that anything worth doing in statistics probably needs to be done because the answer is far from obvious. This means that we must make assumptions and simplifications—a model—that are not entirely realistic but are close enough to give us useful insights. Stangl and Berry expand on this point ([243], p.6).

The names given to these options are also not consistently used. You may see common effect described as "fixed effect (singular)" or as "equal effect".

In a Bayesian context, some of these distinctions become irrelevant, because of the flexible way in which we use probability. In particular, fixed effects and random effects models will lead to different results, but the choice between them is a philosophical one. In practical terms, a Bayesian fixed effects model is simply a Bayesian random effects model with a flat (or wide and uniform) heterogeneity prior distribution. For this reason, we do not consider fixed effects as a distinct class of meta-analysis in this book.

4.1 Heterogeneity in the Data-Generating Process

Meta-analysis often combines different studies, done in slightly different ways, in different places and times. Naturally, this causes some differences in the statistics that they report. Sometimes, we can understand the reason for such a difference, but sometimes we can't, and the best we can do is to model it statistically. It is called *heterogeneity*, from Greek meaning "coming from different sources".

In the previous section, we presented common-effect meta-analyses, which assume that all the variance between studies can be attributed to sampling.

The alternative, called *random effects models*, include heterogeneity as an additional variance, which has scattered the studies' populations' intervention effects (θ_j) around an underlying* intervention effect (θ), and then each study has sampled from their own populations to obtain their reported statistics ($\hat{\theta}_j$).

This means that the observed variance has two components: the sampling within each study (which the reported standard errors estimate), and the differences between the studies (which we meta-analysts must estimate).

$$\begin{aligned} \theta_j &\sim \mathrm{N}(\theta, \tau) \quad ,\forall j \\ \hat{\theta}_j &\sim \mathrm{N}(\theta_j, \widehat{SE}(\hat{\theta}_j)) \quad ,\forall j \end{aligned} \tag{4.1}$$

There are a few aspects that you should pause to consider:

1. In every study ($\forall j$), participants were drawn from a population where the mean intervention effect is θ_j.
2. The study-specific (or local) populations' mean intervention effects are scattered around an underlying (or global) intervention effect, θ. We assume a normal distribution for this with standard deviation τ. This is the heterogeneity distribution.
3. Equivalently, we could instead represent heterogeneity as adding or subtracting some value u_j from the underlying intervention effect θ. This replaces θ_j with $\theta + u_j$, and then $u_j \sim \mathrm{N}(0, \tau)$. Because the u_js are centred on 0, some are positive and others negative: some study populations have higher mean intervention effects than the underlying θ, other are lower.
4. This heterogeneity is treated as a purely random process. In other words, we are acting as though we understand nothing about why the studies differ.
5. The individual studies' estimates of their populations' mean intervention effects, $\hat{\theta}_j$, are in turn drawn from the sampling distribution around their respective θ_js.
6. The mean has a normal sampling distribution, as long as the samples are not too small and the data not too far from normality.
7. The standard error of the mean intervention effect is itself an estimate—hence the large hat on $\widehat{SE}(\hat{\theta}_j)$—though in the first models that follow, we ignore this.

Heterogeneity adds an inter-study variance, $\tau^2 = V(\theta_j)$ to the intra-study variance (standard error squared) of the sampling distribution, $\widehat{SE}(\hat{\theta}_j)^2 = \hat{V}(\hat{\theta}_j)$. Because the two are uncorrelated, the total variance is just the sum of the two component variances (see Equation 1.15). In fact, this is not a practical formula because there are multiple values of

*Names for this are all tricky. Sometimes people say "global" effect, but that implies validity for all people everywhere.

$\widehat{SE}(\hat{\theta}_j)$, one for each study. In the next section, we will show how this concept translates to a statistical model that leads to estimates and inference.

The challenge here is that differences between studies manifest as a sum of two variances: if we observe some estimate of the total variance, and wish to partition this, but know neither of the components, then we cannot proceed.

The DerSimonian-Laird method [53] tackles this in the simplest way, by providing a formula to estimate τ^2, which assumes that the standard errors reported by the studies are correct. (That is to say, $\widehat{SE}(\hat{\theta}_j) = SE(\hat{\theta}_j), \forall j.$) Other methods have been proposed too[24, 220], which take different approaches to including the uncertainty in the standard errors. We will create Bayesian models in this chapter that mirror the principles of the DerSimonian-Laird and another method, the Sidik-Jonkman.

If τ^2 is large, then a more complex model that accounts for heterogeneity is required. This typically means a random effects meta-analysis. Various measures have been proposed to assess the size of τ^2 relative to the total variance. The most common are a chi-squared statistic called Cochran's Q, with accompanying hypothesis test, and Higgins' I^2, which estimates the percentage of total variance arising from heterogeneity.

A common misunderstanding is that one should fit either common effect or random effects, depending on which fits the data best. In fact, the decision should be based on information outside the statistics, about the study populations, interventions, controls and outcomes, and other aspects of study design. Some people contend that heterogeneity should always be included, as studies are not (usually) intended as direct replications of one another, and always have some differences.

You should also be wary of over-optimism about what can be learnt from heterogeneity. Popular ideas such as "personalised medicine", in the Bayesian context, draw on the idea that we can make inferences about individual narrow zones of the posterior distribution pertaining to particular patients. The combined intervention effect θ is informed by all studies in your evidence base, but to make inferences in the tails of the heterogeneity distribution will draw on perhaps only one study, and also have relatively few posterior draws to inform it [28].

4.1.1 DGP for continuous outcomes

We will now amend the data-generating process (DGP) for continuous outcomes, seen for common effects meta-analysis in Section 3.5.1. We will consider only the difference between arms, the mean difference, and in Chapter 8, we look into the possibility of extending this to the individual arms.

As before, participants (i) are randomly drawn from normally distributed populations, but now, the studies (j) do not all draw from the same population. We simplify this by assuming that the studies are themselves drawn from a distribution (of potential studies that might be done). Further, we assume that studies draw from populations with different means, but all have one common population standard deviation, σ. There are only two arms: the intervention ($k = \textbf{Int}$) and the control ($k = \textbf{Ctl}$).

$$\text{Input parameters: } \mu, \theta, \sigma, \tau \tag{4.2}$$

$$\text{Data generation:}$$

$$\theta_j \sim \mathrm{N}(\theta, \tau) \tag{4.3}$$

$$\mu_{j\textbf{Ctl}} = \mu \tag{4.4}$$

$$\mu_{j\textbf{Int}} = \mu + \theta_j \tag{4.5}$$

$$y_{ijk} \sim \mathrm{N}(\mu_{jk}, \sigma) \quad , \forall (i, j, k) \tag{4.6}$$

Taking each line in turn, we can relate them to what we have already covered so far.

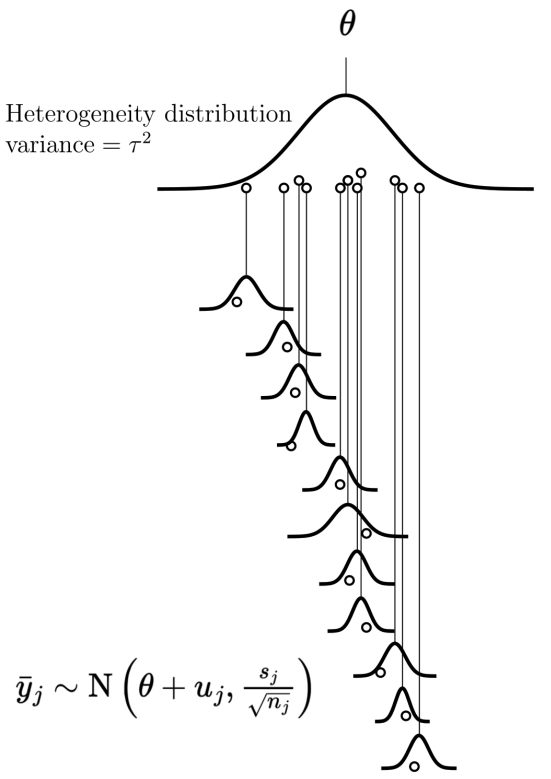

FIGURE 4.1
A continuous outcome data-generating process with heterogeneity. Each study has an observed mean drawn from the relevant normal distribution on the left. We assume the standard errors are perfectly known.

Equation 4.2: there are four parameters, whose values are not known to us, but we will estimate them and the uncertainty around our estimates: the mean outcome in control groups, which we assume is shared by all studies (μ), the population difference between intervention and control arms (θ), the common population standard deviation (σ), and the inter-study standard deviation of heterogeneity (τ).

Equation 4.3: each study has its own population (we will call it *study-specific*) intervention effect, θ_j. This can also be written as:

$$\begin{aligned} u_j &\sim \mathrm{N}(0, \tau) \\ \theta_j &= \theta + u_j \\ \therefore \theta_j &\sim \mathrm{N}(\theta, \tau) \\ \theta_j &= \mu_{j\mathbf{Int}} - \mu_{j\mathbf{Ctl}} \end{aligned}$$

This is the difference between the arm-specific means ($\mu_{j\mathbf{Int}}$ and $\mu_{j\mathbf{Ctl}}$). As a linear combination of two normally distributed variables, we can use Equation 1.15 to find its distribution. Here, the heterogeneity distribution is derived; can you see why its standard deviation, τ, is the same as the standard deviation of the u_js?

Note that the heterogeneity models the scatter of studies' θ_js, not the μ_{jk}s; that is to say, heterogeneity affects the intervention effect (difference between arm means), not the arm means themselves.

Equation 4.4: the study populations for all control arms ($k = 2$) have means equal to μ.

Equation 4.5: the study populations in the green tea arms ($k = 1$) have means equal to $\mu + \theta_j$, so each study will differ somewhat.

Equation 4.6: the individual data in each study and arm come from a normally distributed population with the relevant arm-specific mean and common standard deviation σ.

Now, we can easily simulate data from this, if we input the four parameters, along with n_{j1} and n_{j2}. For example, simulating one study in R:

```
mu <- 0
theta <- (-1)
sigma <- 2
tau <- 0.5

n_1Int <- 100
n_1Ctl <- 120

# heterogeneity:
theta_1 <- rnorm(1, theta, tau)

# arm-specific population means:
mu_1Int <- mu + theta_1
mu_1Ctl <- mu

# data:
y_1Int <- rnorm(n_1Int, mu_1Int, sigma)
y_1Ctl <- rnorm(n_1Ctl, mu_1Ctl, sigma)

# summary stats:
mean(y_1Int)
sd(y_1Int)
mean(y_1Ctl)
sd(y_1Ctl)
t.test(y_1Int, y_1Ctl)
```

On the website, we provide scripts that repeatedly do this simulation (for several studies), and then run a DerSimonian-Laird meta-analysis on the resulting statistics. This allows you to see that DerSimonian-Laird works well if the assumptions are met.

Let's consider next the statistics that each study will provide, to complete our journey back to meta-analysis. Equations 4.7 to 4.10 present the same ideas of sampling that we have already covered.

The observed mean in any arm is drawn from a sampling distribution around the true arm-specific mean:

$$\bar{y}_{jk} \sim \mathrm{N}\left(\mu_{jk}, \frac{\sigma}{\sqrt{n_{jk}}}\right) \quad ,\forall(j,k) \tag{4.7}$$

The variances of each arm's mean can be added together because, as a randomised study, they are uncorrelated. Each study's estimate of the intervention effect is the difference between the observed arm-specific means:

$$\hat{\theta}_j = \bar{y}_{j1} - \bar{y}_{j2} \tag{4.8}$$

Each study's estimate of the intervention effect is distributed about the study-specific mean intervention effect θ_j by a sampling distribution:

$$\hat{\theta}_j \sim \mathrm{N}\left(\theta_j, \sqrt{\frac{\sigma^2}{n_{j1}} + \frac{\sigma^2}{n_{j2}}}\right) \tag{4.9}$$

This is not very useful, because we have no idea what the values of θ_j are, so in Equation 4.10, we link this back further to θ, combining the heterogeneity and sampling distributions (because they too are uncorrelated):

$$\begin{aligned} \therefore \hat{\theta}_j &\sim \mathrm{N}\left(\theta, \sqrt{\frac{\sigma^2}{n_{j1}} + \frac{\sigma^2}{n_{j2}} + \tau^2}\right) \\ \hat{\theta}_j &\sim \mathrm{N}\left(\theta, \widehat{SE}(\hat{\theta}_j)\right) \end{aligned} \tag{4.10}$$

We can write Equation 4.10 in a different way, by introducing u_j, which can be positive or negative with mean 0:

$$\begin{aligned} u_j &\sim \mathrm{N}(0, \tau) \\ \hat{\theta}_j &\sim \mathrm{N}(\theta + u_j, \widehat{\mathrm{SE}}(\hat{\theta}_j)) \end{aligned} \tag{4.11}$$

In BUGS, it will look something like this:

```
u[j] ~ dnorm(0, tau_precision)
theta[j] <- theta + u[j]
theta_hat[j] ~ dnorm(theta[j], se_precision[j])
```

We can easily simulate some study statistics that will look like these, and play with some different values of τ, to get a closer understanding of the model. On the website, we provide code where you supply τ and obtain a forest plot representing simulated study results. This will help you to acquire an instinctive recognition of what different levels of heterogeneity look like, and when the various statistics might be misleading.

This brings us full circle to Equation 4.1. We have started from the data-generating process for individual participant data and seen how this leads to the heterogeneity variance and the sampling error. This two-level scattering is captured in a random effects meta-analysis. It is also an example of what statisticians call a *hierarchical* or *multilevel model*. Hierarchical models are very widely used, for example in cluster-randomised clinical trials, or data from electronic health records, where patients attend one of a range of local health care providers, and we would expect the patients at one location to be somehow different to those at another location.

Now, consider the likelihood methods in Section 2.1, and how they could be applied here—just at an informal level of detail. If we have been given $\hat{\theta}_j$ and $s_{j\bullet}$ or something else that allows you to calculate $\widehat{SE}(\hat{\theta}_j)$, can you apply the equations above to evaluate the likelihood for various values of θ and σ? How might you also estimate τ? Do you need μ for this?

You might feel that this DGP approach is excessively complex for a meta-analysis, especially as we explained that non-Bayesian methods are elaborations on a weighted average. We have spelt out much more detail than you will need in order to perform a meta-analysis so that all the connections between participant data, population parameters and study statistics are clear at this early stage.

The DGP approach has two important benefits. We have to be explicit about all our assumptions, and can no longer hide behind black box analyses. This helps us to question the evidence base and to be prepared to justify all our modelling choices. Also, when we move to Bayesian analyses, we have the framework ready. We can easily adapt the DGP, for example, to allow for non-normal heterogeneity.

4.2 Bayesian Models and Priors

Like we did for common effect meta-analysis in Section 3.6, we will now show the corresponding models, and later the code, to implement a simple random effects meta-analysis. We will use a variety of priors here and in the code that follows, to illustrate options; they are not recommended in general.

4.2.1 Priors

To fit a Bayesian meta-analysis to our data, we need to supply prior distributions for the unknown parameters of the model: $f(\theta)$, $f(\tau)$, $f(u_j)$ in the case of the model in Equation 4.11. Noting the various approaches to prior distributions in Chapter 2, we will use weakly informative priors in this section, to focus attention on the construction of the model.

In experimental studies, it is typical to appeal to equipoise when designing the recruitment and seeking ethical approval. On this basis, a meta-analysis of these experimental studies (such as randomised controlled trials (RCTs)), with weakly informative or diffuse $f(\theta)$, can reasonably use priors with a median of 0: there is equal probability of the intervention effect being on either side.

We have been dealing with mean differences so far, and these can take both negative and positive values, so $f(\theta)$ should be defined for all real numbers from $\theta = -\infty$ to $\theta = +\infty$.

The normal distribution is the usual starting point. t and Cauchy distributions allow higher probability of being further from the mean—and in so doing, allow the computer to get into potentially troublesome territory, where posterior densities are extremely low and it takes a very long time to return to the high density region.

Remember that there is no safe "default" prior that can be applied to all instances of a certain statistic. Prior predictive checking, as described in Chapter 2, will help you detect mis-matches between your priors and your evidence base.

As with any standard deviation, τ can only take positive values, so its priors should do so too. It has been common practice for many years to use inverse-gamma distributions for variances (see Section 2.2.3.2), though this has been criticised in recent years in favour of normal, Cauchy or t-distributions, truncated to only the part above zero[85]. These are often called "half-normal", "half-Cauchy", and so on. We will write them like this: $\tau \sim \mathrm{N}^{+}(0, 2)$. Examples appear in the software-specific sections below. Several alternatives have been suggested in the past[247].

Some people use uniform priors for standard deviations or precisions, as there is usually no prior information or opinion to shape them other than ruling out completely unbelievable

values. However, there is a risk in providing a flat region in the priors or likelihood for some sampling algorithms, and until you are a confident Bayesian modeller, we recommend something like a half-normal so that there is always a gradient† to guide the sampling algorithm back from extreme values.

The prior for each of the u_j parameters is the same, and is the heterogeneity distribution. If you make this normal, with standard deviation τ, then that is the prior: $u_j \sim \mathrm{N}(0, \tau)$.

Note that u_j has a prior, the standard deviation of which (τ) has its own prior! Some people use the term "hyperprior" in this setting, but we prefer not to single it out as acting in any different way to other probabilistic statements in our model (also to avoid the need for hyperhyperpriors further down the line). It simply reflects the probability of finding different values of τ. It also functions to pool information (people often say "borrow information") across all the studies, because they all tell us something about τ.

4.2.2 Empirical priors to update a previous meta-analysis

The main unknown that you will seek to estimate is likely something that serves as an intervention effect between two arms. For this, whether it is a mean difference, a log odds ratio, or log rate ratio, the prior and posterior are going to be approximately normal. Take the green tea meta-analysis as an example ("Analysis 1.4", p.61 [136]). They reported a mean difference of -0.47 units of body mass index (kg/m^2), favouring green tea, and a 95% confidence interval from -0.77 to -0.17. Bearing in mind that the normal 95% confidence interval extends 1.96 standard errors on either side of the mean, that implies that the standard error was $0.3/1.96 = 0.15$. So we can set a prior as $\mathrm{N}(-0.47, 0.15)$, and update it with just the latest studies.

There are some complications to this though. Firstly, our model should be comparable to the one that we are adopting from the previous meta-analysis. If we include heterogeneity, so should they. Inclusion / exclusion criteria should be the same for studies too, and if there is a marked change in the population or study designs and conduct between the old and new studies, then we should not take the old results as a prior for the new.

Secondly, there are other unknowns, notably the heterogeneity standard deviation. We should choose whether to impose a prior on those that is also based on previous results [264], or some kind of diffuse or weakly informative prior. If the posterior is correlated between our intervention effect and some of these other unknowns, then we will perhaps bias our results somewhat by having a more diffuse prior on the other unknown, and allowing it to accommodate less likely values.

This problematic correlation should not occur for simple meta-analyses, but there is no guarantee of that for more complicated models, such as we will construct in Part 3 of this book. In those models, a better idea would be to adopt the study statistics from the old evidence base and include them in your meta-analysis, effectively analysing all the data together from scratch. Unfortunately, as we will see, this can extend well beyond the usual d_{jk} and n_{jk}, or n_{jk}, $\bar{y}_{jk}$, and s_{jk}, and so it could be a lot of work to extract the additional statistics needed for a complicated model.

In summary, although these empirical priors using previous results sound like a good idea, mainly as a labour-saving device, they are not always easy to apply in a way that we can be comfortable will not inadvertently distort the analysis.

---

†... in theory at least; digital rounding error means that there may be *de facto* flat regions of parameter space when the proposed parameter values have either a very small prior density or very low likelihood. Sensible initial values will help us avoid them.

Index