Introduction to the `metap` package

Michael Dewey

July 12, 2020

1 Introduction

1.1 What is this document for?

This document describes some methods for the meta–analysis of p–values (significance values) and their implementation in the package `metap`. I welcome feedback about sources of published examples against which I can test the code and any other comments about either the documentation or the code.

The problem of meta–analysis of p–values is of course not completely unconnected with the more general issue of simultaneous statistical inference.

1.2 Why and when to meta–analyse significance values

The canonical way to meta–analyse a number of primary studies uses estimates of effect sizes from each of them. There are a large number of packages for this purpose available from CRAN and described in the task view `http://CRAN.R-project.org/view=MetaAnalysis`. However sometimes the only available information may be p–values especially when some of the primary studies were published a long time ago or were published in sources which were less rigorous about insisting on effect sizes. The methods outlined here are designed for this eventuality. The situation may also arise that some of the studies can be combined in a conventional meta–analysis using effect sizes but there are many others which cannot and in that case the conventional meta–analysis of the subset of studies which do have effect sizes may usefully be supplemented by an overall analysis of the p–values.

Just for the avoidance of doubt, if each study has produced a proportion and the goal is to synthesise them to a common estimate or analyse the differences between them then the standard methods are appropriate not the ones outlined here. The p–values in this document are significance levels.

The methods are referred to by the name of the function in `metap`. Table 1 shows other descriptions of each method.

Function name	Description(s)	
	Eponym	
`invchisq`	Lancaster's method	Inverse chi square
`invt`		Inverse t
`logitp`		Logistic
`meanp`		
`meanz`		
`maximump`		
`minimump`	Tippett's method	
`sumlog`	Fisher's method	Chi square (2 df)
`sump`	Edgington's method	Uniform
`sumz`	Stouffer's method	Normal
`truncated`	Truncated Fisher	rank–truncated
`votep`		
`wilkinsonp`	Wilkinson's method	

Table 1: Methods considered in this document

2 Preparation for meta–analysis of p–values

2.1 Preliminaries

I assume you have installed R and `metap`. You then need to load the package.

```
> library(metap)
```

2.2 Directionality

It is usual to have a directional hypothesis, for instance that treatment is better than control. For the methods described here a necessary preliminary is to ensure that all the p–values refer to the same directional hypothesis. If

the value from the primary study is two–sided it needs to be converted. This is not simply a matter of halving the quoted p–value as values in the opposite direction need to be reversed. A convenience function `two2one` is provided for this.

```
> pvals <- c(0.1, 0.1, 0.9, 0.9, 0.9, 0.9)
> istwo <- c(TRUE,  FALSE, TRUE, FALSE, TRUE, FALSE)
> toinvert <- c(FALSE, TRUE, FALSE, FALSE, TRUE, TRUE)
> two2one(pvals, two = istwo, invert = toinvert)

[1] 0.05 0.90 0.45 0.90 0.55 0.10
```

Note in particular the way in which 0.9 is converted under the different scenarios.

2.3 Plotting

```
> print(validity)

 [1] 0.015223 0.005117 0.224837 0.000669 0.004063 0.549106 0.052925 0.024674
 [9] 0.004618 0.287803 0.738475 0.009563 0.071971 0.000003 0.001040 0.031221
[17] 0.005274 0.098791 0.067441 0.250210
```

It would be a wise precaution to examine the p–values graphically or otherwise before subjecting them to further analysis. Two functions are provided for this purpose: `plotp` and `schweder`.

2.3.1 Plotting using `plotp`

The `plotp` provides a Q–Q plot of the p–values to detect departure from the uniform distribution. An example is shown in Figure 1. The standard line through through the quartiles is superimposed. This is the function which is called when the `plot` method is used on an object returned by any of the meta–analysis functions.

```
> plotp(validity)
```

2.3.2 Plotting using `schweder`

A function `schweder` provides plots with a variety of informative lines superimposed. It plots the ordered p–values, $p_{[i]}$, against i. Although the original motivation for the plot is Schweder and Spjøtvoll (1982) the function uses
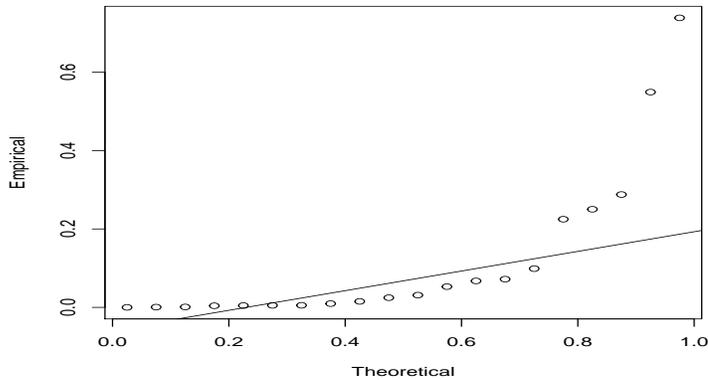
Figure 1: Q–Q plot from `plotp`

a different choice of axes due to Benjamini and Hochberg (2000). We will use an example dataset on the validity of student ratings quoted in Becker (1994). Figure 2a shows the plot from `schweder` which is the same as from `plotp` but without the line.

```
> schweder(validity)
```

`schweder` also offers the possibility of drawing one of a number of straight line summaries. The three possible straight line summaries are shown in Figure 2b and are:

- the lowest slope line of Benjaimin and Hochberg which is drawn by default as solid,

- a least squares line drawn passing through the point $k + 1, 1$ and using a specified fraction of the points which is drawn by default as dotted,

- a line with user specified intercept and slope which is drawn by default as dashed.

```
> schweder(validity, drawline = c("bh", "ls", "ab"),
+    ls.control = list(frac = 0.5), ab.control = list(a = 0, b = 0.01))
```

2.4   Reporting problems in the primary studies

Another issue is what to do with studies which have simply reported on whether a conventional level of significance like 0.05 was achieved or not. If the exact associated p cannot be derived from the statistics quoted in the primary source then the value of the level achieved, in this case 0.05, can

4

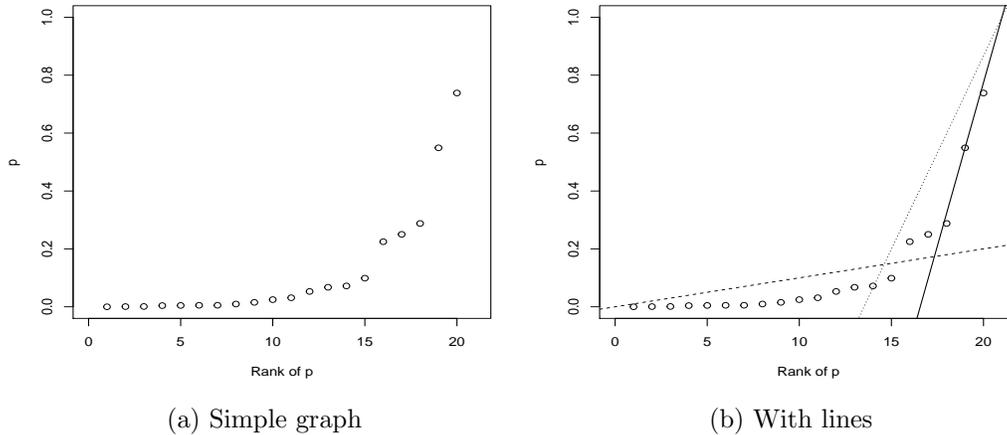(a) Simple graph   (b) With lines

Figure 2: Output from schweder

be used although this may be conservative. Studies which simply report not significant could be included as having $p = 1$ (or $p = 0.5$ if it is known that the direction was right) although this is very conservative. The theory of handling p–values which have been truncated like this has been developed by Zaykin et al. (2002) and `truncated` provides a convenience wrapper for two methods available in other CRAN packages.

3 Using the methods

All the methods in the package take as their first argument the vector of p–values. To use Fisher's method as an example:

```
> sumlog(validity)

chisq =  159.82  with df =  40  p =  2.989819e-16
```

A few require extra information. Those which rely on inverse transformations often need a vector of degrees of freedom. Currently this applies to `invchisq` and `invt`. Stouffer's method in `sumz` optionally uses weights if a vector of weights is provided.

4 Miscellanea

Extractor functions The standard `print` and `plot` methods are provided.

Reading An annotated bibliography is provided by Cousins (2008)

References

B J Becker. Cambining significance levels. In H Cooper and L V Hedges, editors, *A handbook of research synthesis*, chapter 15, pages 215–235. Russell Sage, New York, 1994.

Y Benjamini and Y Hochberg. On the adaptive control of the false disovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25:60–83, 2000.

R D Cousins. Annotated bibliography of some papers on combining significances or p–values, 2008. arXiv:0705.2209.

T Schweder and E Spjøtvoll. Plots of p–values to evaluate many tests simultaneously. *Biometrika*, 69:493–502, 1982.

D V Zaykin, L A Zhivotovsky, P H Westfall, and B S Weir. Truncated product method for combining p–values. *Genetic Epidemiology*, 22:170–185, 2002.