Plotting in the `metap` package

Michael Dewey

December 18, 2025

1 Introduction

1.1 What is this document for?

This document describes how and why to plot p–values in the `metap` package. Examining the p–values graphically or otherwise before subjecting them to further analysis is useful to provide a visual impression of their distribution and to check for excess p–values at both extremes.. Three functions are provided for this purpose: `albatros`, `plotp`, and `schweder`.

1.2 Example datasets

As our example we use various data-sets:

teachexpect Effect of teacher expectations on student IQ (Becker, 1994)

validity The validity of student ratings of their instructors (Becker, 1994).

zhang The effect of the timing of exercise interventions for patients with cardiovascular disease (Zhang et al., 2016)

```
> library(metap)
> data(dat.metap)
> teach <- dat.metap$teachexpect
> validity <- dat.metap$validity$p
> zhang <- dat.metap$zhang
> 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
```

2 Plotting using `plotp`

The `plotp` provides a Q–Q plot of the p–values to detect departure from the uniform distribution.

```
> plotp(validity, main = "Validity data")
```


(a) Q–Q plot from `plotp`
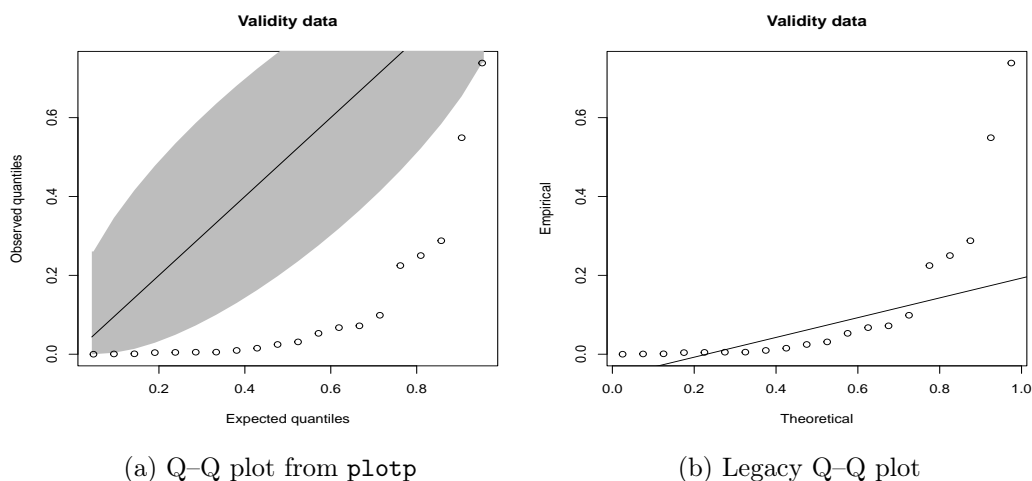
(b) Legacy Q–Q plot

Figure 1: Plots of validity data

Figure 1a shows the resulting plot. The line represents a fit to the uniform distribution and the polygon is a simultaneous confidence region such that if any point lies outside it we reject the null hypothesis that the points are drawn iid from a uniform. Small p–values are to the left of the plot

The format of plot shown in Figure 1a was first introduced in version 1.8 of `metap`. The previous plotting function is still available and it is possible to produce this plot by setting the `plotversion` parameter to "old" in the call to `plotp`. An example is shown in Figure 1b which first calls `sumlog`. The legacy one will always remain an option.

```
> plotp(validity, main = "Validity data", plotversion = "old")
```

Note that the `plot` method for objects of class `"metap"` uses the new version of the plot. This change was introduced in version 1.9 of this package.

There are many possible options which can be passed to the plotting function and hence to the `qqconf` plotting routine. The documentation for the `qqconf` package should be consulted for details. The `qqconf` package vignette is also very helpful. We will look at one of those options here though.

```
> plotp(teach)

> plotp(teach, log10 = TRUE)
```
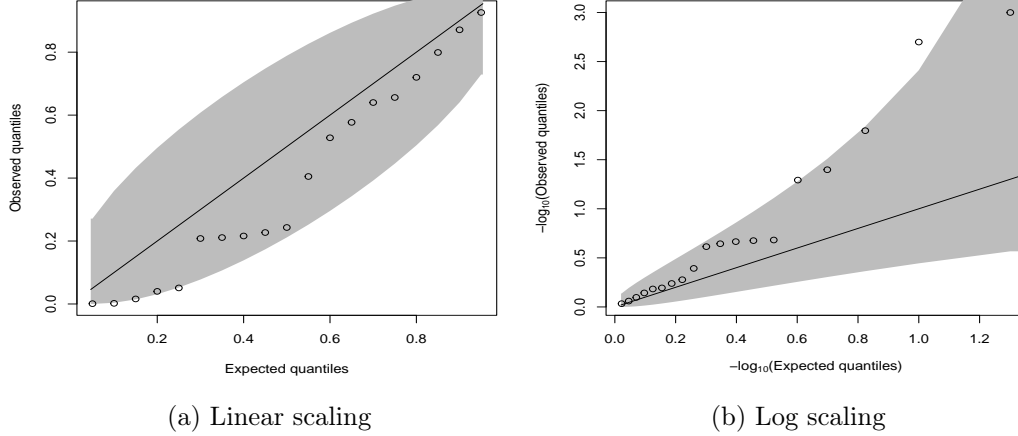

(a) Linear scaling   (b) Log scaling

Figure 2: Teacher expectancy data

Figure 2 shows the teacher expectancy data using the default scaling in sub–figure **??**. It is hard to see whether some of the points fall outside the boundary. However if we use the log–scaling option shown in sub–figure **??** it becomes much clearer. Note that the scale is reversed between the sub–plots and in **??** the small p–values are now on the right. So the cluster of points near the bottom left of the sub–figure **??** are hard to distinguish as to whether they lie inside the boundary or not. In the log scaling of sub–figure **??** where they appear towards the top right it is much clearer that one does fall outside the boundary and two others are borderline. This reflects the fact that for most of the methods in the `metap` package the overall p–value is below 0.05. For instance using the logit method we have

```
> logitp(teach)

t =  2.763487  with df =  99  p =  0.003409951
```

3   Plotting using `schweder`

A function `schweder` provides plots with a variety of informative lines superimposed. It plots the ordered p–values, $p_{[i]} : p_{[1]} \leq \ldots p_{[2]} \leq \ldots p_{[i]} \ldots \leq p_{[k-1]} \leq p_{[k]}$, against i. Although the original motivation for the plot is

Schweder and Spjøtvoll (1982) the function uses 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 3a shows the plot from `schweder`.

```
> 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 3b 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))
```
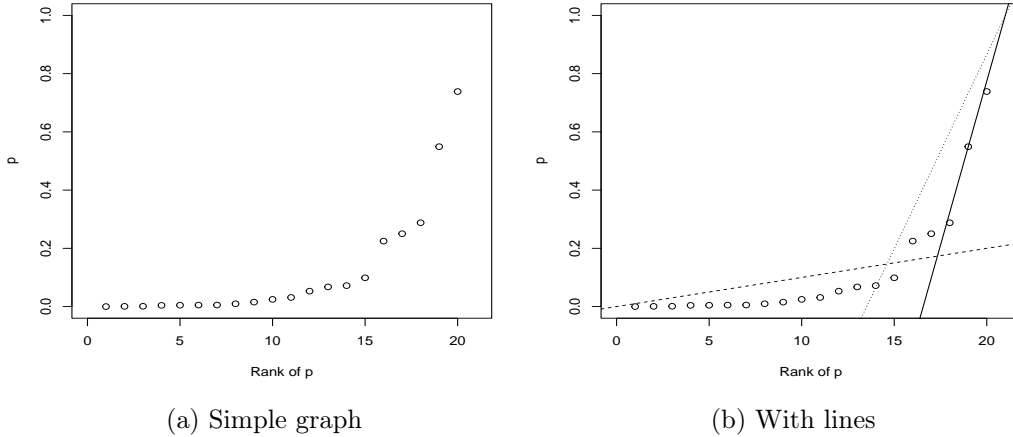

(a) Simple graph   (b) With lines

Figure 3: Output from schweder

4   The albatros plot

The albatros plot was introduced in Harrison et al. (2017) which should be consulted for more details. Basically it consists of plotting a possibly

transformed sample size against the transformed p–values. The default is to use $\sqrt{N}$ for the y–axis and a log transformation for the x–axis. The scale for the y-axis is user selectable. The original scale in the Stata version is $(\log_{10} n)^2$ which is obtained by setting `yscale` to "classic". For small N the default of "sqrt" is very similar to "classic". The plot also contains contours of constant effect size. A number of possible options are available for effect size type: correlation, standardised mean difference, and odds ratio.

```
> validity <- dat.metap$validity
> fit.v <- albatros(validity$p, validity$n,
+     contours = list(type = "corr", contvals = c(0.25, 0.5, 0.8), ltys = 1:3),
+         axes = list(ylimit = c(1,200),  lefttext = "Negative correlation",
+             righttext = "Positive correlation"),
+     main = "Validity")
```
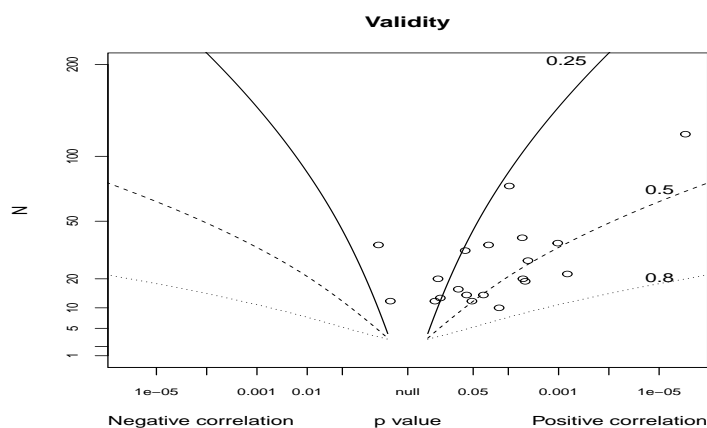

Figure 4: Albatros plot from of the validity data

Figure 4 shows the result. Most of the points clearly correspond to positive and substantial correlations although a few are in the opposite direction although not far from the null p–value (0.5).

Of course if the actual effect sizes are available it would be better to use one of the conventional methods for meta–analysing them. Harrison et al. (2017) outline possible use cases for this method even so.

If the studies come from different groups one might use meta–regression with a moderator for group membership if one had the effect sizes. In the absence of effect sizes the albatros plot can display the points using different symbols for groups. This would enable a visual check on whether the groups differed.

```
> data(dat.metap)
> zhang <- dat.metap$zhang
```

```
> fit.z <- albatros(zhang$p, zhang$n,
+    contours = list(type = "smd", contvals = c(0.25, 0.5, 1), ltys = 1:3),
+    plotpars = list(pchs = letters[unclass(zhang$phase)]),
+    axes = list(lefttext = "Favours control", righttext = "Favours exercise"),
+    main = "Zhang"
+ )
> legend(-4, 11, c("Acute", "Healing", "Healed"),
+    pch = c("a", "b", "c"), bg = "white"
+ )
```
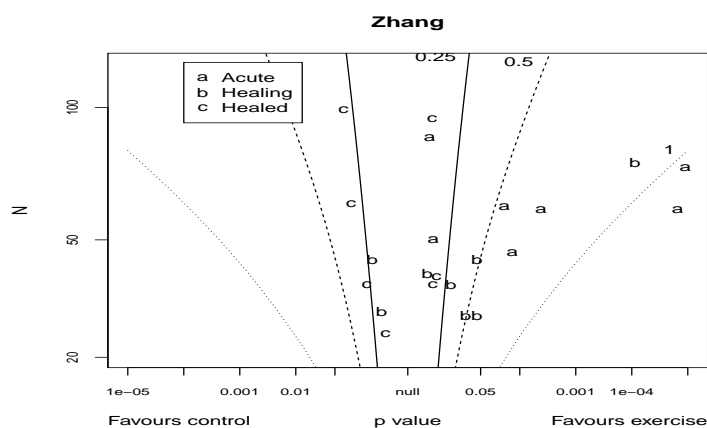
Figure 5: Albatros plot of the Zhang et al data

Figure 5 shows an example using the Zhang et al data-set. The studies involved come from three groups corresponding to three different periods of initiation of exercise. The points are labelled accordingly: "a" initiation during the acute phase, "b" during the healing phase and "c" during the healed phase. The difference between the groups is quite clear here. In fact in Zhang et al. (2016) the results are handled with stratification into three separate analyses and meta–regression was not used.

If some studies had given effect sizes but others did not then an albatros plot with the points marked for group membership and with appropriate contour lines would provide a visual check on whether the unavailable effect sizes were similar to the available ones.

References

B J Becker. Combining 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.

S Harrison, H E Jones, R M Martin, S J Lewis, and J P T Higgins. The albatros plot: A novel graphical tool for presenting the results of diversely reported studies in a systematic review. *Research Synthesis Methods*, 8: 281–289, 2017.

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

Y-M Zhang, Y Lu, D Yang, H-F Wu, Z-P Bian, J-D Xu, C-R Gu, L-S Wang, and X-J Chen. The effects of different initiation time of exercise training on left ventricular remodeling and cardiopulmonary rehabilitation in patients with left ventricular dysfunction after myocardial infarction. *Disability and Rehabilitation*, 38:268–276, 2016.