O'Connell-Dobson-Schouten estimators of agreement applied to data from Landis and Koch (1976)

Mark Clements
Karolinska Institutet

Abstract

This vignette applies the O'Connell-Dobson-Schouten estimators of agreement to a classical set of data on inter-rater agreement from Landis and Koch (1976). The analysis here follows the analysis given in O'Connell and Dobson (Biometrics 1984; 40: 973–983) and Schouten (Statistica Neerlandica 1982: 36: 45-61). We propose that the O'Connell-Dobson-Schouten estimators are valuable and encourage their broader use.

Keywords: agreement.

O'Connell and Dobson (1984) and Schouten (1982) provide a technical discussion on the statistical properties of averaged measures of agreement, particularly un-weighted and weighted kappa statistics that adjust for the probability of chance agreement. Such measures of agreement are common for studies of multi-rater agreement with nominal or ordinal variables. We have adapted the Fortran code from the 1984 paper by O'Connell and Dobson for use in R and implemented the algorithms described by Schouten (1982) in Fortran. In the following, we apply the resulting package to the dataset analysed in both articles. We propose that the O'Connell-Dobson-Schouten estimators are valuable and encourage their broader use.

Landis and Koch (Biometrics 1977; 33: 363–374) provide a dataset on inter-rater agreement by seven pathologists for classifying carcinoma in situ for uterine cancer. The data are included in the **oconnell** package as the `landis` data, which is a matrix.

After loading the package, we can produce summary statistics for `i=1`, which is the unweighted analysis, for the linear weights (`i=2`) or for quadratic weights (`i=3`). The summary statistics include marginal summaries and $\hat{S}_{av}$ for each slide.

```
> require(oconnell)
> ## Table 1 (O'Connell and Dobson, 1984)
> summary(fit <- oconnell(landis, weight="unweighted"))

O'Connell-Dobson estimator (unweighted)

Sav(hetero):        0.361290 (se: 0.028881; 95% CI: 0.306811, 0.419587)
Sav(homoge):        0.354335 (se: 0.030018; 95% CI: 0.297924, 0.415112)
Pr(Overall agreement due to chance | hetero):       < 2.22e-16
Pr(Overall agreement due to chance | homoge):       < 2.22e-16
```

```
Observed marginal distributions for categories:

         1          2          3          4          5
0.28087167 0.25423729 0.36440678 0.07384988 0.02663438

Observed marginal distributions for categories by observer:

          1         2         3           4           5
A 0.2203390 0.2203390 0.3220339 0.186440678 0.050847458
B 0.2288136 0.1016949 0.5847458 0.059322034 0.025423729
C 0.2627119 0.3559322 0.3135593 0.050847458 0.016949153
D 0.3220339 0.4067797 0.1949153 0.067796610 0.008474576
E 0.1355932 0.2627119 0.4491525 0.118644068 0.033898305
F 0.5254237 0.2627119 0.1694915 0.008474576 0.033898305
G 0.2711864 0.1694915 0.5169492 0.025423729 0.016949153

Agreement statistics S_i for the individual items:

          1           2           3           4           5           6
 0.08088076  1.00000000  1.00000000  0.34348626  1.00000000  0.34348626
          7           8           9          10          11          12
 0.60609175  0.21218351  0.27783488  0.60609175  0.60609175  0.60609175
         13          15          16          17          18          19
 0.60609175  0.21218351  0.08088076  0.60609175  0.21218351  0.21218351
         22          23          24          25          26          27
 0.08088076  0.60609175  0.34348626  0.34348626  1.00000000  0.34348626
         28          29          30          31          32          33
 0.08088076  0.34348626  0.60609175  1.00000000  0.27783488  1.00000000
         34          35          36          37          38          39
 1.00000000  0.27783488  0.27783488  0.08088076  0.01522939  0.34348626
         40          41          42          43          44          45
 0.08088076  0.60609175  1.00000000  0.08088076  0.27783488  0.60609175
         46          47          48          49          51          52
-0.05042199  0.21218351  0.34348626  0.08088076  0.60609175  0.27783488
         53          54          55          56          57          58
 0.27783488  0.01522939  0.60609175  0.60609175  0.01522939  1.00000000
         59          60          61          62          63          64
 1.00000000  0.60609175  0.08088076  0.27783488  0.08088076  0.21218351
         65          66          67          68          69          70
 0.60609175  0.08088076  1.00000000  0.34348626  0.27783488  1.00000000
         71          72          73          74          76          77
 0.60609175  0.27783488  0.60609175  0.01522939  0.60609175  0.34348626
         78          79          80          81          82          83
 0.08088076  0.34348626 -0.11607336  1.00000000  0.21218351  0.08088076
         84          85          86          87          88          89
 0.27783488 -0.18172474  0.60609175  0.60609175  0.01522939 -0.11607336
```

```
        90          91          92          93          94          95
0.08088076  0.01522939 -0.11607336  0.21218351  0.34348626  0.27783488
        96          98          99         100         101         102
-0.11607336  0.21218351  0.21218351  0.08088076  0.34348626  0.34348626
       103         104         105         106         107         108
1.00000000  0.01522939  0.60609175  0.08088076  0.21218351  0.08088076
       110         111         112         113         114         115
0.21218351  0.60609175  0.21218351  0.08088076  0.08088076  0.21218351
       116         117         118         119         120         121
0.60609175  0.34348626  0.08088076  0.60609175  1.00000000  0.21218351
       122         123         124         126
-0.11607336 -0.11607336  0.60609175  0.01522939
```

A simple print of the object provides a short description of the estimator. This is shown here for the linear and quadratic weights.

```
> require(oconnell)
> ## Table 1 (O'Connell and Dobson, 1984), continued
> print(update(fit, weight="linear"))

O'Connell-Dobson estimator (linear weights)

Sav(hetero):         0.515924 (se: 0.034694; 95% CI: 0.448050, 0.583216)
Sav(homoge):         0.509672 (se: 0.036048; 95% CI: 0.439295, 0.579667)
Pr(Overall agreement due to chance | hetero):        < 2.22e-16
Pr(Overall agreement due to chance | homoge):        < 2.22e-16

> print(update(fit, weight="quadratic"))

O'Connell-Dobson estimator (quadratic weights)

Sav(hetero):         0.646884 (se: 0.039399; 95% CI: 0.566433, 0.719791)
Sav(homoge):         0.641728 (se: 0.040832; 95% CI: 0.558428, 0.717270)
Pr(Overall agreement due to chance | hetero):        < 2.22e-16
Pr(Overall agreement due to chance | homoge):        < 2.22e-16
```

Table 3 of O'Connell and Dobson (1984) includes an analysis where the slides are grouped by level of disagreement. We reproduce the table here.

```
> slideTypeGroups <-
+     list(c(2,3,5,26,31,34,42,58,59,67,70,81,103,120),
+          c(7,10:13,17,23,30,41,51,55,56,60,65,71,73,76,86,87,105,111,116,119,124),
+          c(4,6,24,25,27,29,39,48,68,77,79,94,101,102,117),
+          c(9,32,36,44,52,62,84,95),
+          c(35,53,69,72),
+          c(8,15,18,19,47,64,82,93,98,99,107,110,112,115,121),
```

```
+           c(1,16,22,49,63,66,78,90,100,113),
+           c(28,37,40,61,108,114,118),
+           106,
+           43,
+           83,
+           c(54,57,88,91,126),
+           c(74,104),
+           38,
+           46,
+           c(89,122),
+           c(80,92,96,123),
+           85)
```

The average $\hat{S}_i$ in those groups can be readily calculated by:

```
> data.frame(SlideType=1:18,
+             S1=sapply(slideTypeGroups,
+                 function(ids) mean(fit$s1[as.character(ids)])),
+             S2=sapply(slideTypeGroups,
+                 function(ids) mean(fit$s2[as.character(ids)])))

   SlideType          S1           S2
1          1  1.00000000  1.000000000
2          2  0.60609175  0.601802471
3          3  0.34348626  0.336337452
4          4  0.27783488  0.269971197
5          5  0.27783488  0.269971197
6          6  0.21218351  0.203604942
7          7  0.08088076  0.070872432
8          8  0.08088076  0.070872432
9          9  0.08088076  0.070872432
10        10  0.08088076  0.070872432
11        11  0.08088076  0.070872432
12        12  0.01522939  0.004506178
13        13  0.01522939  0.004506178
14        14  0.01522939  0.004506178
15        15 -0.05042199 -0.061860077
16        16 -0.11607336 -0.128226332
17        17 -0.11607336 -0.128226332
18        18 -0.18172474 -0.194592587
```

which follows part of Table 2. Finally, Table 5 from O'Connell and Dobson (1984) can be easily reproduced by dichotomising the outcomes:

```
> oconnell(landis==1)

O'Connell-Dobson estimator (unweighted)
```

```
Sav(hetero):        0.563058 (se: 0.043851; 95% CI: 0.476079, 0.646324)
Sav(homoge):        0.558453 (se: 0.045153; 95% CI: 0.469036, 0.644234)
Pr(Overall agreement due to chance | hetero):       < 2.22e-16
Pr(Overall agreement due to chance | homoge):       < 2.22e-16

> oconnell(landis==2)

O'Connell-Dobson estimator (unweighted)

Sav(hetero):        0.159782 (se: 0.029600; 95% CI: 0.109879, 0.226580)
Sav(homoge):        0.152886 (se: 0.030545; 95% CI: 0.102073, 0.222719)
Pr(Overall agreement due to chance | hetero):       7.1863e-14
Pr(Overall agreement due to chance | homoge):       3.8278e-11

> oconnell(landis==3)

O'Connell-Dobson estimator (unweighted)

Sav(hetero):        0.373709 (se: 0.036869; 95% CI: 0.304686, 0.448289)
Sav(homoge):        0.364046 (se: 0.038646; 95% CI: 0.292140, 0.442584)
Pr(Overall agreement due to chance | hetero):       < 2.22e-16
Pr(Overall agreement due to chance | homoge):       < 2.22e-16

> oconnell(landis==4)

O'Connell-Dobson estimator (unweighted)

Sav(hetero):        0.180271 (se: 0.043233; 95% CI: 0.110276, 0.280679)
Sav(homoge):        0.173971 (se: 0.043852; 95% CI: 0.103787, 0.276947)
Pr(Overall agreement due to chance | hetero):       < 2.22e-16
Pr(Overall agreement due to chance | homoge):       1.8569e-14

> oconnell(landis==5)

O'Connell-Dobson estimator (unweighted)

Sav(hetero):        0.626828 (se: 0.145957; 95% CI: 0.330852, 0.850890)
Sav(homoge):        0.626413 (se: 0.146334; 95% CI: 0.329882, 0.850997)
Pr(Overall agreement due to chance | hetero):       < 2.22e-16
Pr(Overall agreement due to chance | homoge):       < 2.22e-16
```

We can also compare the raters using the results due to Schouten (1982).

```
> fit2 <- schouten(landis)
> summary(fit2)
```

```
Schouten estimator (unweighted)

Average kappa:        0.361290 (se: 0.028881; 95% CI: 0.306811, 0.419587)
Pr(Overall agreement due to chance):        < 2.22e-16

Observed marginal distributions for categories by observer:

          1         2         3           4           5
A 0.2203390 0.2203390 0.3220339 0.186440678 0.050847458
B 0.2288136 0.1016949 0.5847458 0.059322034 0.025423729
C 0.2627119 0.3559322 0.3135593 0.050847458 0.016949153
D 0.3220339 0.4067797 0.1949153 0.067796610 0.008474576
E 0.1355932 0.2627119 0.4491525 0.118644068 0.033898305
F 0.5254237 0.2627119 0.1694915 0.008474576 0.033898305
G 0.2711864 0.1694915 0.5169492 0.025423729 0.016949153

Agreement statistics for each rater:

    Kappa [Lower, Upper] Pr(kappa_av=kappa_rater)
A 0.37274 0.30394 0.4471                  0.60718
B 0.40591 0.33691 0.4788                  0.03444 *
C 0.38173 0.31292 0.4556                  0.40306
D 0.33866 0.27026 0.4145                  0.33175
E 0.32894 0.25801 0.4086                  0.23888
F 0.24269 0.17536 0.3257                  < 1e-05 ***
G 0.46538 0.40385 0.5280                  < 1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Affiliation:

Mark Clements
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Email: mark.clements@ki.se