Measurement error model for correlation coefficient estimation

Beiying Ding
Robert Gentleman

May 1, 2024

Introduction

The `MeasurementError.cor` package fits a two-stage measurement error model for estimating correlation between two random variables under bivariate normality. It's application is perhaps most relevant for the gene expression data where both point and standard estimates are available. We have shown that the proposed measurement error corrected correlation estimate has lower bias compared with the usual sample pearson correlation. For details, refer to Ding and Gentleman (2003) as well as R help pages associated with each function.

The `cor.me.vector` and `cor.me.matrix` functions

The `cor.me.vector` calculates the measurement error model estimate of correlation between two observed vectors whereas `cor.me.matrix` calculates all pairwise measurement error model estimate of correlation in the matrix.

```
> library(MeasurementError.cor)
> exp <- matrix(abs(rnorm(100,1000,20)),ncol=10)
> se <- matrix(abs(rnorm(100,50,5)),ncol=10)
> cor.me.vector(exp[1,],se[1,],exp[2,],se[2,])

$estimate
      corr.me     corr.true           mu1           mu2            s1
  -0.98440585   -0.52915982 1005.37818233  988.84283877   22.84994789
```

```
          s2
   0.01106919

$counts
function gradient
     107      100

$convergence
[1] 1

> cor.me.matrix(exp,se)

$corr.true
             [,1]       [,2]       [,3]        [,4]       [,5]       [,6]
 [1,]  1.00000000 -0.5291598  0.9113568 -0.52446473  0.7440874  0.9906254
 [2,] -0.52915982  1.0000000  0.6852980 -0.23305223  0.7473190  0.8815792
 [3,]  0.91135679  0.6852980  1.0000000 -0.19393133 -0.5729944 -0.6888922
 [4,] -0.52446473 -0.2330522 -0.1939313  1.00000000  0.5723113 -0.4742381
 [5,]  0.74408738  0.7473190 -0.5729944  0.57231133  1.0000000  0.9764163
 [6,]  0.99062544  0.8815792 -0.6888922 -0.47423806  0.9764163  1.0000000
 [7,]  0.06476678  0.7324308 -0.7945068  0.05003489  0.7461199  0.7560410
 [8,]  0.64102319 -0.6690578  0.3873492 -0.85269943 -0.6099240  0.8174638
 [9,] -0.38012738 -0.7137460 -0.7432988  0.86050247 -0.1162776 -0.7728238
[10,]  0.71366231  0.4481993 -0.7683728 -0.21320764  0.9109250  0.8225879
             [,7]       [,8]       [,9]      [,10]
 [1,]  0.06476678  0.6410232 -0.3801274  0.7136623
 [2,]  0.73243081 -0.6690578 -0.7137460  0.4481993
 [3,] -0.79450675  0.3873492 -0.7432988 -0.7683728
 [4,]  0.05003489 -0.8526994  0.8605025 -0.2132076
 [5,]  0.74611993 -0.6099240 -0.1162776  0.9109250
 [6,]  0.75604103  0.8174638 -0.7728238  0.8225879
 [7,]  1.00000000 -0.8452887  0.8236808  0.7842034
 [8,] -0.84528873  1.0000000  0.8052799  0.8313397
 [9,]  0.82368082  0.8052799  1.0000000  0.5031295
[10,]  0.78420342  0.8313397  0.5031295  1.0000000

>
```

the quantity of interest, i.e. the model estimate of the correlation between the true value of two random variables whereas `cor.me` is the model estimate of correlation between the measurement errors of the two random

variables. The second quantity may not be of interest. `mu1`,`mu2` and `s1`, `s2` are the estimated mean and standard deviation of the two random variables. `cor.me.matrix` only returns the estimated correlation matrix.

References

Beiying Ding and Robert Gentleman. Measurement error model for correlation coefficient estimation and its application in microarray analysis. 2003.