Disattenuation by simulation

An estimation and inferential tool for correlation amidst estimable measurement error

Michael A. Lawrence, Dalhousie University — mike.lawrence@dal.ca

Introduction

Traditional calculations of correlation and corresponding inferential confidence intervals assume that variables of interest are measured without error, a condition that rarely holds in the real world.

Spearman (1904) noted that measurement error serves to decrease the observed correlation estimate (**r_{obs}**) relative to the true correlation (**ρ**). Furthermore, the present work demonstrates that in the context of measurement error, some traditional inferential approaches to correlation become overly liberal

The DS method

The present work, dubbed “Disattenuation by simulation” (**DS**), seeks to provide a means by which the effects of error may be quantified and overcome.

One common approach to the problem of measurement error is the application of repeated measurement. In addition to reducing the effect of measurement error on parameter estimates of interest, this procedure also provides opportunity for the estimation of measurement error itself.

The DS method begins by obtaining such error estimates (**λ_{obs}**) within each unit of observation and each variable of interest. With this information, we may ask:

*Given **λ_{obs}**, what **ρ** is most likely to have produced **r_{obs}**?*

We obtain likelihood estimates of **r_{obs}** for each of a number of candidate values for **ρ** (**ρ'**) by repeatedly generating data using **ρ'** and **λ_{obs}**, then observing the correlation (**r_{sim}**) of the resultant simulated data.

Each **ρ'** therefore becomes associated with a distribution of **r_{sim}** values within which the likelihood and the percentile of **r_{obs}** can be calculated, facilitating the generation of correlation estimates and inferential intervals, respectively.

Monte Carlo validation

Performance comparisons of traditional correlation procedures (**Raw**) and the **DS** method were achieved via Monte Carlo simulation in which the methods were simultaneously applied to a large number of data sets for which **ρ** was known and to which measurement error was added.

To create the data sets, 4 variables were manipulated:

ρ – the true correlation
λ – measurement noise
N – # of units of observation
k – # of observations made within each unit of observation

For each combination of chosen levels of the above variables, 2000 simulated data sets were generated. After applying **Raw** and **DS** methods to each, three performance measures were calculated, displayed at right (for simplicity of presentation, effects of **N** and **k** are collapsed).

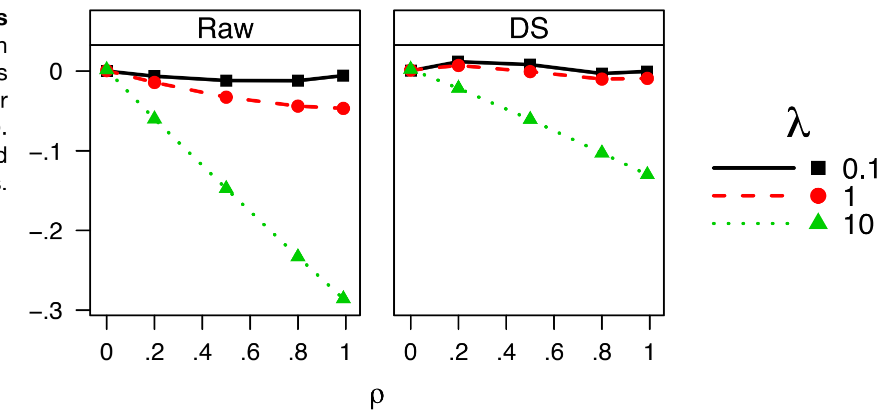
Conclusions

The Monte Carlo results validate that when compared to traditional estimation and inferential procedures the **DS** method achieves more accurate confidence intervals and estimates of **ρ** while losing no power to reject the null hypothesis of zero correlation.

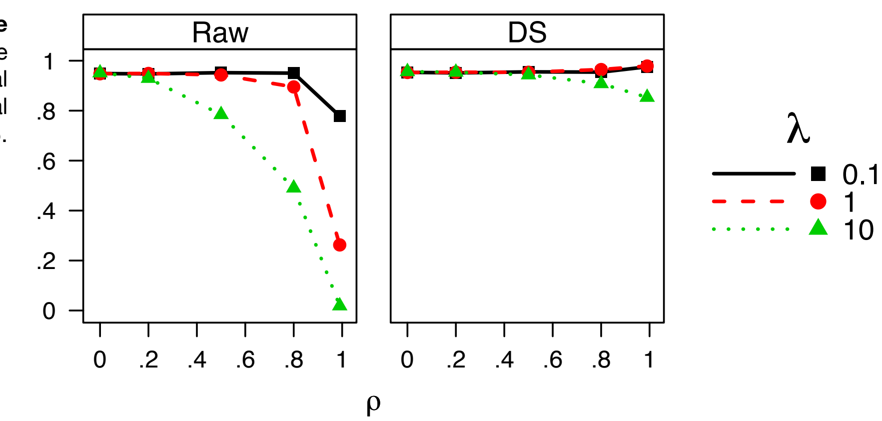
The **DS** method is especially superior in the context of strong true relationships between noisy variables, suggesting that its use may be of particular benefit to applications that seek the comparison of such relationships (e.g. factor analysis).

Finally, while other corrections to the problem of attenuation exist (Spearman, 1904; Liu et al, 1978), further Monte Carlo comparisons (omitted here for brevity) show that the **DS** method outperforms these competitors as well.

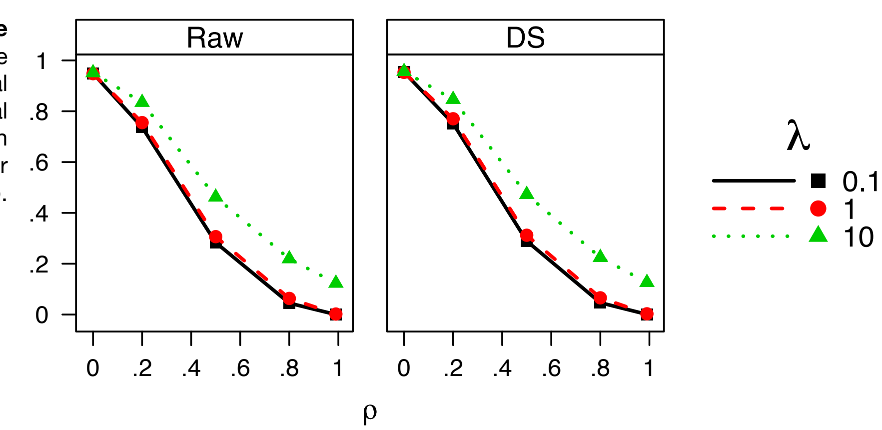
Bias
The degree to which correlation estimates differed from their respective value of **ρ**. Ideal performance would contain zero bias.



ρ capture rate
The rate at which the 95% confidence interval captured **ρ**. Ideal performance is .95.



Zero capture rate
The rate at which the 95% confidence interval captured zero. Ideal performance is .95 when **ρ** is zero, 0 for all other values of **ρ**.



Acknowledgements
This work was supported by a Julie Payette NSERC Research Scholarship and achieved using high performance computing facilities made freely available by the Shared Hierarchical Academic Research Computing Network.