How to calcuate SMD for latency and proportion data

Shinichi Nakagawa, modif by ML

20/07/2018"

Background and Context

We would like to conduct a meta-analysis of cognitive bias (e.g. whether some condition make animals more optimistic or pessimistic). In these cognitive bias studies, people measure the latency to respond to stimuli (latency) or the proportion of time responded (proportion.

Now we want to calculate standardized mean difference (SMD; often known as Cohen's d or Hedges' g) for latency and proportion which are non-normally distributed. In our case, we have latency which appears log-normally distributed and "proportion" which follows binomial distributions.

Original SMD definition

$$\text{SMD} = \frac{\bar{x}_E - \bar{x}_C}{sd_{pool}} \cdot J$$

$$sd_{pool} = \sqrt{\frac{(n_C - 1)sd_C^2 + (n_E - 1)sd_E^2}{n_C + n_E - 2}}$$

$$J = \left(1 - \frac{3}{4(n_C + n_E) - 9}\right)$$

where $\bar{x}_C$ and $\bar{x}_E$ are the means of the control and experimental group, respectively, sd is sample standard deviation (sd^2 is sample variance) and n is sample size. J is to correct bias due to small sample size.

Importantly, we also have its sampling variance

$$se^2_{\text{SMD}} = \frac{n_C + n_E}{n_C n_E} + \frac{\text{SMD}^2}{2(n_C + n_E)}$$

The point estimate and sampling variance of SMD can be obtained by using function in the `metafor` package, called `escalc`. But also, you could write your own functions.

Latency

For the SMD formula, we use $\overline{\ln(x)}$ and $sd^2_{\ln}$ instead of $\bar{x}$ and sd^2. To get these, we can use the delta method for approximating mean and variance on the transformed scale.

The delta method

$$E\left(f(x)\right) \approx f(x) + \frac{\text{var}(x)\left(f''(x)\right)}{2}$$
$$\text{var}\left(f(x)\right) \approx \text{var}(x)^{\cdot}\left(f'(x)\right)^2$$

Prepration: the first and the second derivative

$$\ln(\bar{x})' = \frac{1}{\bar{x}}$$

$$\ln(\bar{x})'' = -\frac{1}{\bar{x}^2}$$

Mean & variance

$$\overline{\ln(x)} = \ln(\bar{x}) - \frac{sd^2}{2\bar{x}^2}$$

$$sd_{\ln}^2 = \frac{sd^2}{\bar{x}^2}$$

Mean & variance: alternative

Alternatively, probably more accurately, if we assume x is log-normally distributed, then, we have exact solutions for the mean and variance on the natural log scale.

$$\overline{\ln(x)} = \ln(\bar{x}) - \ln\left(\sqrt{1 + \frac{sd^2}{\bar{x}^2}}\right)$$

$$sd_{\ln}^2 = \ln\left(1 + \frac{sd^2}{\bar{x}^2}\right)$$

After putting these in the Hedges' g formulas and reversing the sign, we get for latency data (replacing C = Better, E = Worse):

$$\mathrm{H}_g = -\frac{\overline{\ln(x)}_{Better} - \overline{\ln(x)}_{Worse}}{sd_{pool}} \cdot J$$

$$where:$$

$$sd_{pool} = \sqrt{\frac{(n_{Worse} - 1)sd_{\ln(Worse)}^2 + (n_{Better} - 1)sd_{ln(Better)}^2}{n_{Worse} + n_{Better} - 2}}$$

$$J = \left(1 - \frac{3}{4(n_{Worse} + n_{Better}) - 9}\right)$$

and its variance:

$$se_{\mathrm{H}_g}^2 = \frac{n_{Better} + n_{Worse}}{n_{Better} \cdot n_{Worse}} + \frac{\mathrm{H}_g^2}{2(n_{Better} + n_{Worse})}$$

Proportion

Likewise, for the SMD formula, we use $\overline{\mathrm{logit}(x)}$ and sd_{logit}^2 instead of $\bar{x}$ and sd^2. To get these, we can use the delta method for approximating mean and variance on the transformed scale.

Prepration: the first and the second derivative

$$\text{logit}(\bar{x})' = \frac{1}{\bar{x}} + \frac{1}{1-\bar{x}}$$

$$\text{logit}(\bar{x})'' = \frac{1}{(1-\bar{x})^2} - \frac{1}{\bar{x}^2}$$

Mean & variance

$$\overline{\text{logit}(x)} = \text{logit}(\bar{x}) + \frac{sd^2}{2}\left(\frac{1}{(1-\bar{x})^2} - \frac{1}{\bar{x}^2}\right)$$

$$sd^2_{\text{logit}} = sd^2\left(\frac{1}{\bar{x}} + \frac{1}{1-\bar{x}}\right)^2$$

Variance: alternative

Alternatively, since the logit distribution is supposed to map x to a logistic distribution with the mean of 0 and the variance of $\pi^2/3$. Therefore:

$$sd^2_{\text{logit}} = \frac{\pi^2}{3}$$

This formula is useful when sd^2 is not available. However, this does not account for overdispersion due to, for example, individual variation so it is better to use the one above (i.e, when we have sd^2).

After putting these in the Hedges' g formulas we get for proportion data (replacing C = Better, E = Worse):

$$\text{H}_g = \frac{\overline{logit(x)}_{Better} - \overline{logit(x)}_{Worse}}{sd_{pool}} \cdot J$$

$$where:$$

$$sd_{pool} = \sqrt{\frac{(n_{Worse}-1)sd^2_{logit(Worse)} + (n_{Better}-1)sd^2_{logit(Better)}}{n_{Worse}+n_{Better}-2}}$$

$$J = \left(1 - \frac{3}{4(n_{Worse}+n_{Better})-9}\right)$$

and its variance:

$$se^2_{\text{H}_g} = \frac{n_{Better}+n_{Worse}}{n_{Better} \cdot n_{Worse}} + \frac{\text{H}_g^2}{2(n_{Better}+n_{Worse})}$$