The Shieh's d and its relation with Cohen's d

Marie Delacre[1]

[1] Service of Analysis of the Data, Université Libre de Bruxelles, Belgium

Author Note

Correspondence concerning this article should be addressed to Marie Delacre, CP191, avenue F.D. Roosevelt 50, 1050 Bruxelles. E-mail: marie.delacre@ulb.ac.be

Abstract

Keywords: keywords

Word count: X

The Shieh's d and its relation with Cohen's d

Unlike the classical Cohen's δ, Shieh's δ depends on the sample size ratio (that I will call later nratio). For the same amount of differences between two means, same standard deviations (sd) and σ-ratio, Shieh's δ will vary as a function of the nratio. To illustrate the relation between Shieh's δ and the nratio, we can calculate the parameter across a range of nratio. We will first study the Shieh's δ (and its relation with Cohen's δ) when variances are equal between groups. We will then go through the relation when variances are unequal between groups.

When variances are equal between groups

As a first example, in Figure 1, Cohen's δ and Shieh's δ are calculated for different configurations where the observed mean difference ($\mu_1 - \mu_2$) is 1, the total sample size is 200 and standard deviations σ_1 and σ_2 both equals 2. As one can see, as long as variances are equal between groups, Cohen's δ remains constant for all nratio, unlike Shieh's δ. Moreover, Shieh's δ achieves its maximum value when $n_1 = n_2$ (in Figure 1, Shieh's δ equals 0.25 when nratio is 1) [1]. One can also see that at the maximum point value, the relation between Shieh's δ and Cohen's δ is as follows:

$$\delta_{Cohen} = 2 \times max(\delta_{Shieh}) <--> max(\delta_{Shieh}) = \frac{\delta_{Cohen}}{2} \tag{1}$$

When plotting both parameters against the log of the nratio, one can more easily observe that the Shieh's δ departs symmetrically from it's maximum value as long as the nratio moves away from 1 (i.e. when log(nratio) = 0; see Figure 2). The relation between all Shieh's δ values and its value when nratio=1 can be expressed as follows:

[1] Because we keep total sample size (n) constant, the larger n_1, the smaller n_2. As a consequence, when variances are equal between groups, the larger $\frac{\sigma_1^2 \times n}{n_1}$, the smaller $\frac{\sigma_2^2 \times n}{n_2}$. $min(\frac{\sigma_1^2 \times n}{n_1} + \frac{\sigma_2^2 \times n}{n_2})$ will be achieved when $\frac{\sigma_1^2 \times n}{n_1} = \frac{\sigma_2^2 \times n}{n_2}$ and therefore when $n_1 = n_2$.

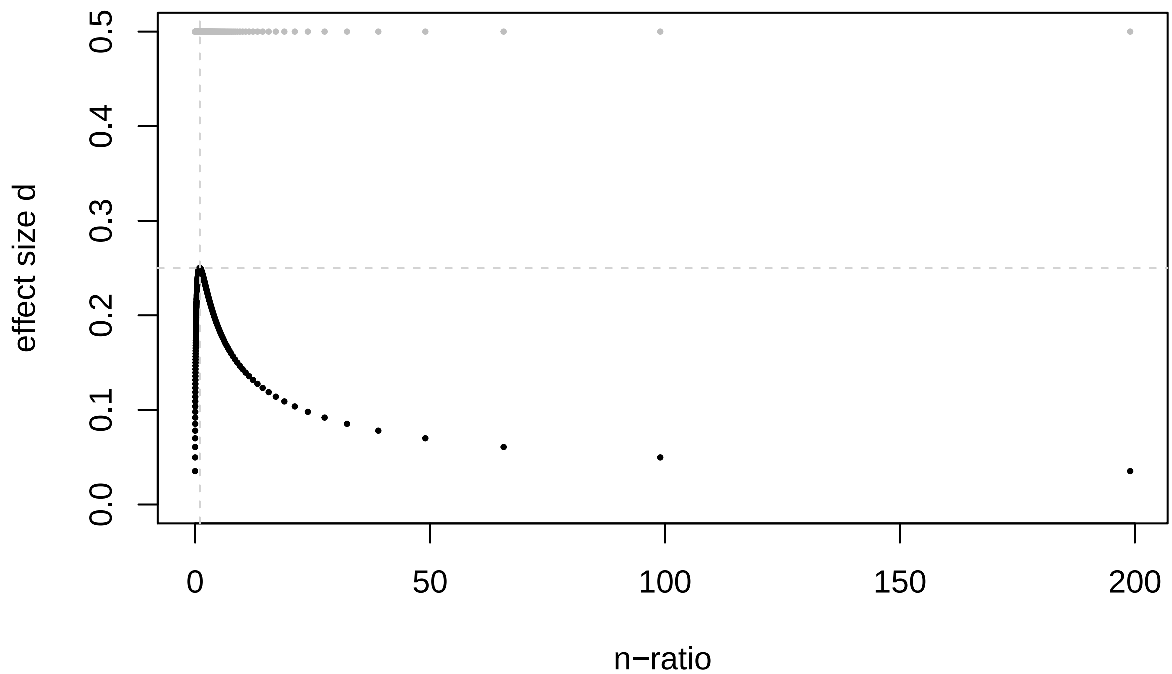

Figure 1. Comparison of Shieh's d (black dots) and Cohen's d (grey dots) when mu1 - mu2 = 1, N = 200 and sigma 1 and sigma 2 both equals 2

$$max(\delta_{Shieh}) = \delta_{Shieh} \times \frac{nratio + 1}{2 \times sqrt(nratio)} \tag{2}$$

31 Because we know from (1) than $max(\delta_{Shieh})$ is half of the value of δ_{Cohen}, we can
32 therefore deduce a more general relation between Cohen's δ and Shieh's δ:

$$\delta_{Cohen} = \delta_{Shieh} \times \frac{nratio + 1}{sqrt(nratio)} \tag{3}$$

33 This relations remains true as long as variances are equal between groups.

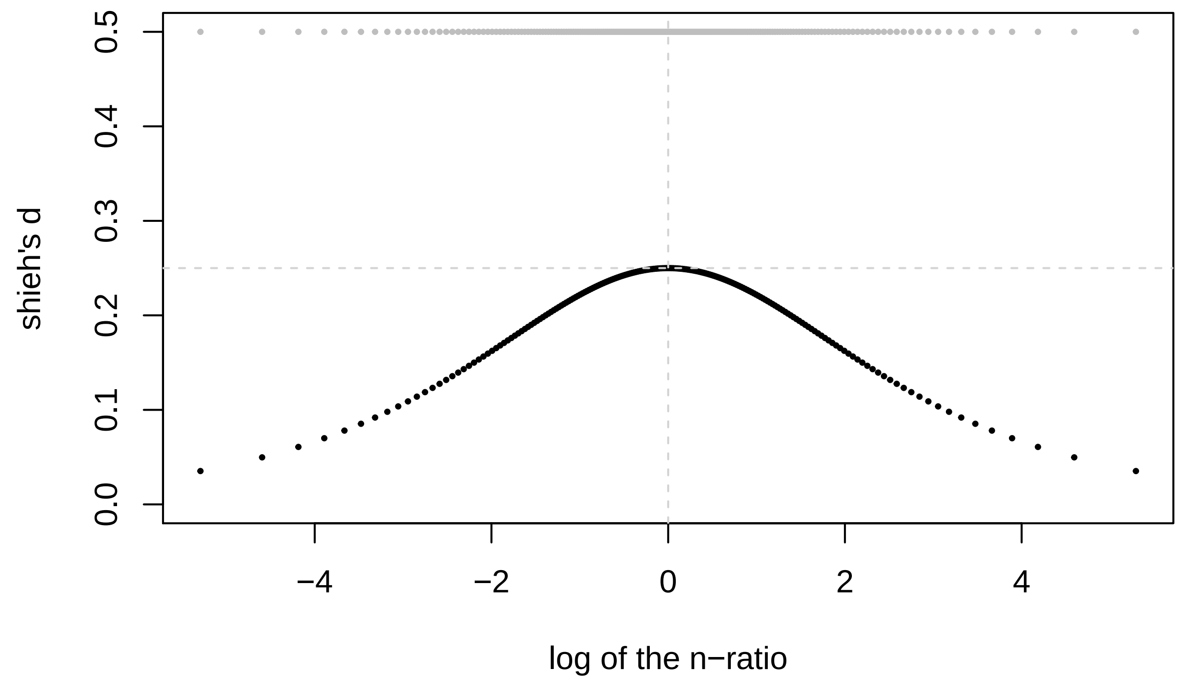

Figure 2. Comparison of Comparison of Shieh's d (black dots) and Cohen's d (grey dots) when mu1 - mu2 = 1, N = 200 and delta 1 = delta 2 = 2

When variances are unequal between groups

In Figure 3, Cohen's δ and Shieh's δ are calculated for different configurations where the observed mean difference ($\mu_1 - \mu_2$) is 1, the total sample size is 200 and standard deviations σ_1 and σ_2 are respectively 4 and 3 (left) or 3 and 4 (right). As one can see, when variances are unequal between groups, Cohen's δ no longer remains constant for all nratio.

Once again, it is easier to study the influence of the nratio on both parameters when plotting them against the log of the nratio, as done in Figure 4. **Cohen's** δ is the difference between both groups means, divided by a pooled error term:

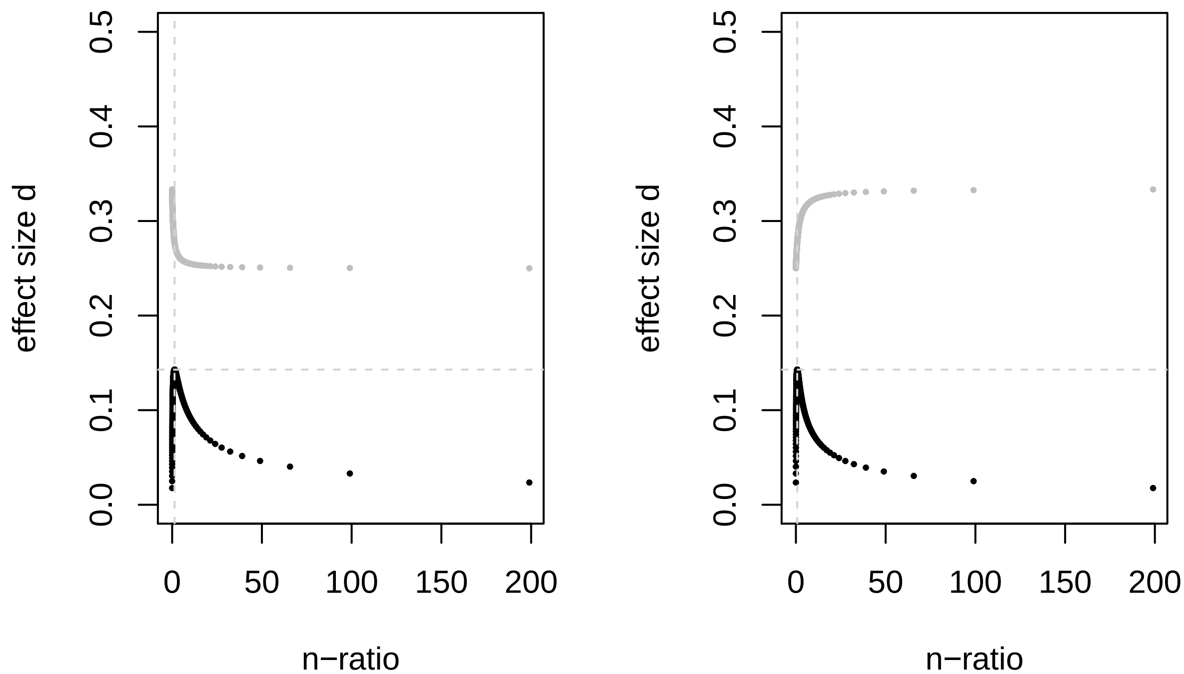

Figure 3. Comparison of Shieh's d (black dots) and Cohen's d (grey dots) when mu1 - mu2 = 1, N = 200 and sigma 1 and sigma 2 are respectively 4 and 2 (left) or 2 and 4 (right)

$$\delta_{Cohen} = \sqrt{\frac{\mu_1 - \mu_2}{\frac{(n_1-1)\times\sigma_1^2+(n_2-1)\times\sigma_2^2}{n_1+n_2-2}}} \tag{4}$$

This implies that all samples are considered as issued from a common population variance (hence the assumption of homoscedasticity). When there is heteroscedasticity, if the larger variance is associated with the larger sample size (i.e. the colored parts on both plots in Figure 4), the error term is overestimated and therefore, the Cohen's δ is decreased. The smallest value is achieved when the sample size of the group associated with the largest variance equals n-1=199. On the other side, if the larger variance is associated with the smaller sample size (i.e. the non-colored parts of both plots), the error term is underestimated and therefore, the Cohen's δ is increased. The largest value is achieved when the sample size of the group associated with the largest variance equals 1.

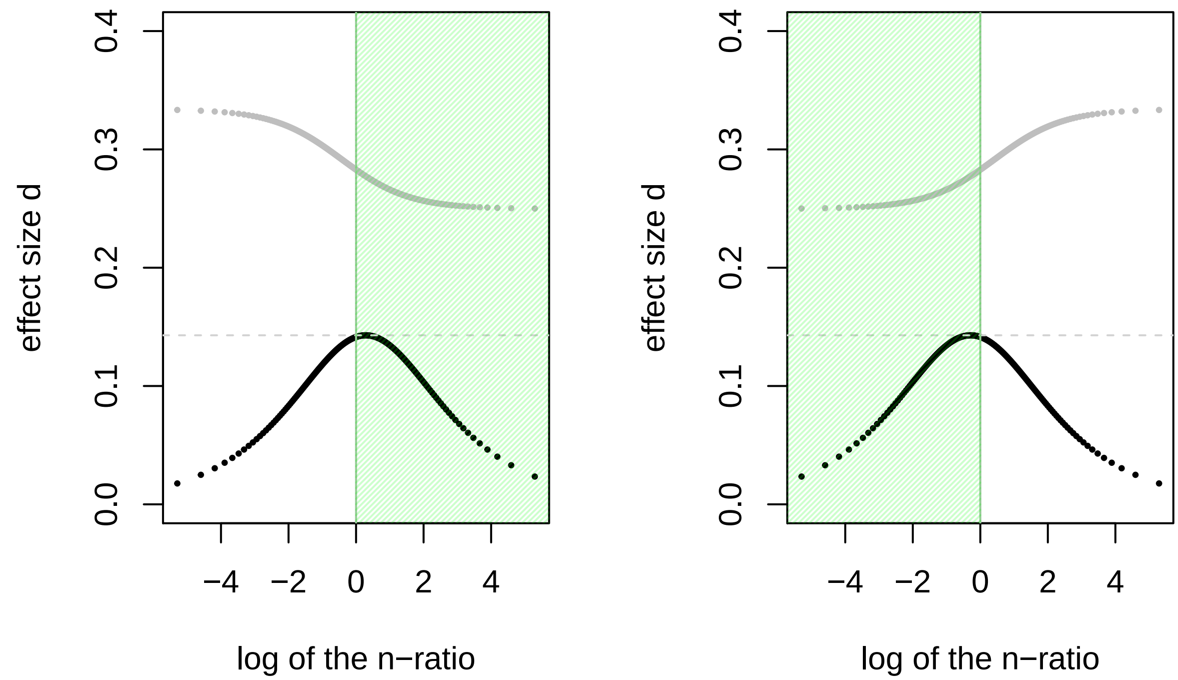

Figure 4. Comparison of Shieh's d (black dots) and Cohen's d (grey dots) when mu1 - mu2 = 1, N = 200 and sigma 1 and sigma 2 are respectively 4 and 2 (left) or 2 and 4 (right)

While it remains true that when $n_1 = n_2$, the Cohen's δ is exactly as twice as large as Shieh's δ, the maximum Shieh's δ value is not achieved when the nratio equals 1 (i.e the log of the nratio equals 0). When there is heteroscedasticity and unequal sample sizes, the maximum Shieh's δ is always achieved when there is a positive correlation between variances and sample sizes and the more unequal the variances, the further from 1 the nratio associated with the maximum. As an illustration, we computed Shieh's δ values for difference configurations where $\sigma_2 = 1$, total sample size = 10000 and $\mu_1 - \mu_2 = 1$, and the nratio varies from $\frac{1}{n-1}$ to n-1, and we extracted the nratio at which Shieh's δ was the largest ($=nratio_{largest\delta}$). All configurations varied as a function of σ_1, which varied from .02 to 50, in steps of .02 (so does the σ-ratio). In Figure 5, we plotted the σ-ratio against $nratio_{largest\delta}$.

As one can observe, when the σ-ratio is larger than 1 (i.e. $\sigma_1 > \sigma_1$), the nratio associated with the largest Shieh's δ value is also larger than 1(i.e. $n_1 > n_2$). On the other side, when the σ-ratio is lower than 1 (i.e. $\sigma_1 < \sigma_1$), the nratio associated with the largest Shieh's δ value is also lower than 1 (i.e. $n_1 < n_2$), and the relation between $nratio_{largest\delta}$ and σ-ratio seems linear.

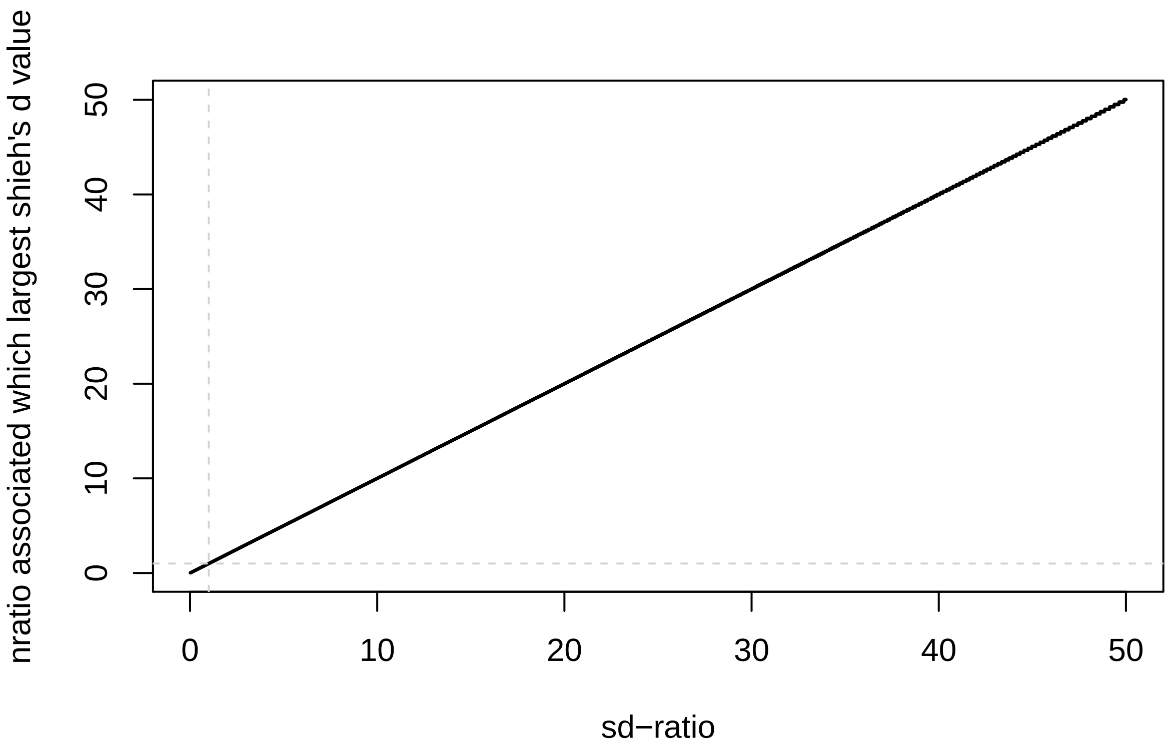

Figure 5. At which nratio the largest value of Shieh's d is achieved, as a function of the sd-ratio (= sd1/sd2)

This result can be easily explained: **Shieh's** δ is the difference between both groups means, divided an error term where variances σ_j are weighted by $\frac{n}{n_j}$:

$$\delta_{Shieh} = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 \times n}{n_1} + \frac{\sigma_2^2 \times n}{n_2}}} \tag{5}$$

It implies that when there is a positive correlation between sample size and σ-ratio, the group with smallest variance is given more weight. As a consequence, the error term is decreased and the Shieh's δ is increased. When there is a negative correlation between sample size and σ-ratio, the group with largest variance is given more weight and therefore, the error term is increased, and the Shieh's δ is decreased. FINISH TO EXPLAIN THIS