Theoretical variance, as a function of population parameters

Delacre Marie[1]

[1] ULB

Theoretical variance, as a function of population parameters

The variance

Cohen's d_s

When variances are equal across populations.

When $\delta_{Cohen} = 0$. When the population effect size is zero, the variance of Cohen's d_s can be simplified as follows:

$$Var_{Cohen's\ d_s} = \frac{N(N-2)}{n_1 n_2 (N-4)}$$

The **variance** of Cohen's d_s is a function of total sample size (N) and the sample size allocation ratio ($\frac{n_1}{n_2}$):

- The larger the total sample size, the lower the variance. The bias tends to zero when the total sample size tends to infinity (see Figure 1)

- The further the sample sizes allocation ratio is from 1, the larger the variance (see Figure 2)

When $\delta_{Cohen} \neq 0$. While the variance of δ_{Cohen} still depends on the total sample size and the sample sizes allocation ratio, it also depends on the population effect size (δ_{Cohen}). The larger the population effect size, the larger the variance. Note that the effect of the population effect size decreases when sample sizes increase, as $\lim_{n \to \infty} \left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} \right)^2 \right] = 0.$

Glass's d_s

When variances are equal across populations.

When $\delta_{Glass} = 0$. When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n_c - 1}{n_c - 3}\left(\frac{1}{n_c} + \frac{1}{n_e}\right)$$

The **variance** of Glass's d_s is a function of the sample sizes of both control and experimental group (n_c) as well as of the sample sizes allocation ratio.

- The larger the sample sizes, the lower the variance (Figure 3)

The sample sizes ratio associated with the lowest variance is not exactly 1 (because of the term $\frac{df}{df-2}$, df depending only on n_c), but is very close to 1 (and the larger the total sample size, the closer to 1 is the sample sizes ratio associated with the lowest variance).And the further from this sample size ratio, the larger the variance (see Figure 4).

When $\delta_{Glass} \neq 0$***.*** While the variance of δ_{Glass} still depends on the total sample size and the sample sizes allocation ratio, it also depends on the population effect size (δ_{Cohen}). The larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the control group increases, as $\lim_{n \to \infty}\left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)}\right)^2\right] = 0$ (see Figure 5).

Note: while the sample sizes ratio associated with the lowest variance was very close to 1 with a null population effect size, this is not true anymore when the population effect size is not zero. Indeed, because of the second term in the adddition, when computing the variance, one gives much more weight to the effect size of the control group (see Figure 6), and the larger the population effect size the truer. For example, when δ_{glass}= 4, the lowest variance will occure when n_c is approximately 3 times larger than n_e. When δ_{glass}= 7, the lowest variance will occure when n_c is approximately 5 times larger than n_e, etc.

When variances are unequal across populations, with equal sample sizes.

When $\delta_{Glass} = 0$***.*** When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n-1}{n(n-3)}\left(1+\frac{\sigma_e^2}{\sigma_c^2}\right)$$

Where $n = N/2$=sample size of each group. The variance is therefore a function of the total sample size and the SD-ratio ($\frac{\sigma_c}{\sigma_e}$).

- The larger the total sample size, the lower the variance (See Figure 7)

- The larger the SD-ratio (i.e. the larger is σ_c in comparison with σ_e), the lower the variance (see Figure 8). However, the effect of the SD-ratio decreases when sample sizes increase, because $\lim_{n\to\infty}\left[\frac{n-1}{n(n-3)}\right] = 0$.

When $\delta_{Glass} \neq 0$**.** While the variance of δ_{Glass} still depends on the total sample size and the SD-ratio, it also depends on the population effect size (δ_{Glass}). The larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the control group increases, as $\lim_{n\to\infty}\left[\frac{df}{df-2}-\left(\frac{\sqrt{\frac{df}{2}}\times\Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)}\right)^2\right] = 0$ (see Figure 5).

When variances are unequal across populations, with unequal sample sizes.

When $\delta_{Glass} = 0$**.** When the population effect size is zero, the variance of Glass's d_s can be simplified as follows:

$$Var_{Glass's\ d_s} = \frac{n_c-1}{n_c-3}\left(\frac{1}{n_c}+\frac{\sigma_e^2}{n_e\sigma_c^2}\right)$$

The variance of Glass's d_s is therefore a function of the total sample size (N), the SD-ratio and the interaction between the sample sizes ratio and the SD-ratio.

- Whatever the SD and sample sizes pairing, increasing n_c and/or n_e will decrease the variance (see Figure 9).

- The effect of the sample sizes ratio depends on the SD-ratio:

- We previously mentioned that when $\sigma_c = \sigma_e$, the variance is minimized when sample sizes of both groups are almost identical (see Figure 4), meaning that it is more efficient, in order to reduce variance, to add subjects uniformly in both groups;
- When $\sigma_e > \sigma_c$, more weight is given to n_e, meaning that it is more efficient, in order to reduce variance, to add subjects in the experiental group (n_e; see Figure 10);
- When $\sigma_c > \sigma_e$, less weight is given to n_e, meaning that it is more efficient, in order to reduce variance, to add sujects in the control group (n_c; see Figure 11).

- Finally, there is also a main effect of the SD-ratio: the larger is σ_c in comparison with σ_e, the lower the variance, as we can observe in Figure 12

Note that the effect of the SD-ratio, and the interaction effect between SD-ratio and sample sizes ratio decreases when the sample size of the control group increases (because $\frac{n_c-1}{n_c-3}$ get closer to 1).

***When* $\delta_{Glass} \neq 0$.** While the variance of δ_{Glass} still depends on the total sample size, the SD-ratio and the interaction between the SD-ratio and the sample sizes ratio, it also depends on the population effect size (δ_{Glass}): the larger the population effect size, the larger the variance. However, the effect of the population effect size decreases when the sample size of the control group increases, as $\lim_{n\to\infty}\left[\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)}\right)^2\right] = 0$

Note: when the population effect size was null, when $\sigma_e > \sigma_c$, it was much more efficient to add subjects in the experimental group in order to reduce the variance (because much more weight were given to n_e). When $\delta_{Glass} \neq 0$, it is important to add subjects in both groups in order to reduce the variance (because $\frac{df}{df-2} - \left(\frac{\sqrt{\frac{df}{2}} \times \Gamma\left(\frac{df-1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)}\right)^2$ is only a function of the sample size of the control group). With huge population effect size, it is even always more important to add subjects in the control group (e.g. when $\delta_{Glass} = 30$).

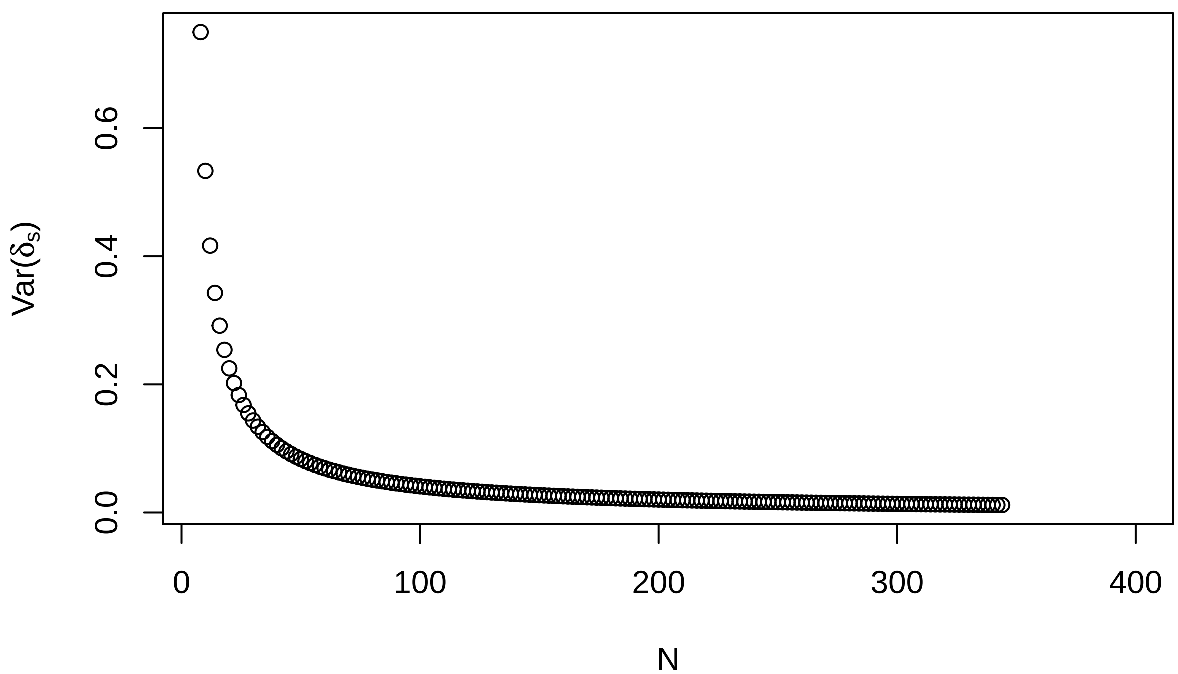


Figure 1. ...

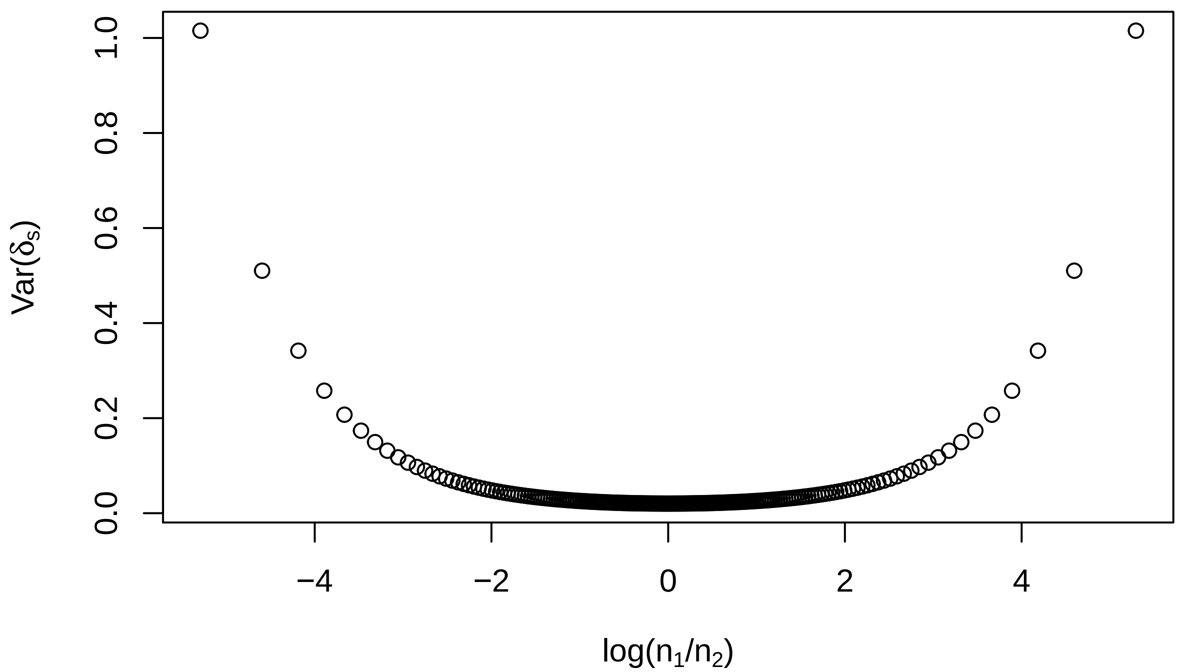


Figure 2. ...

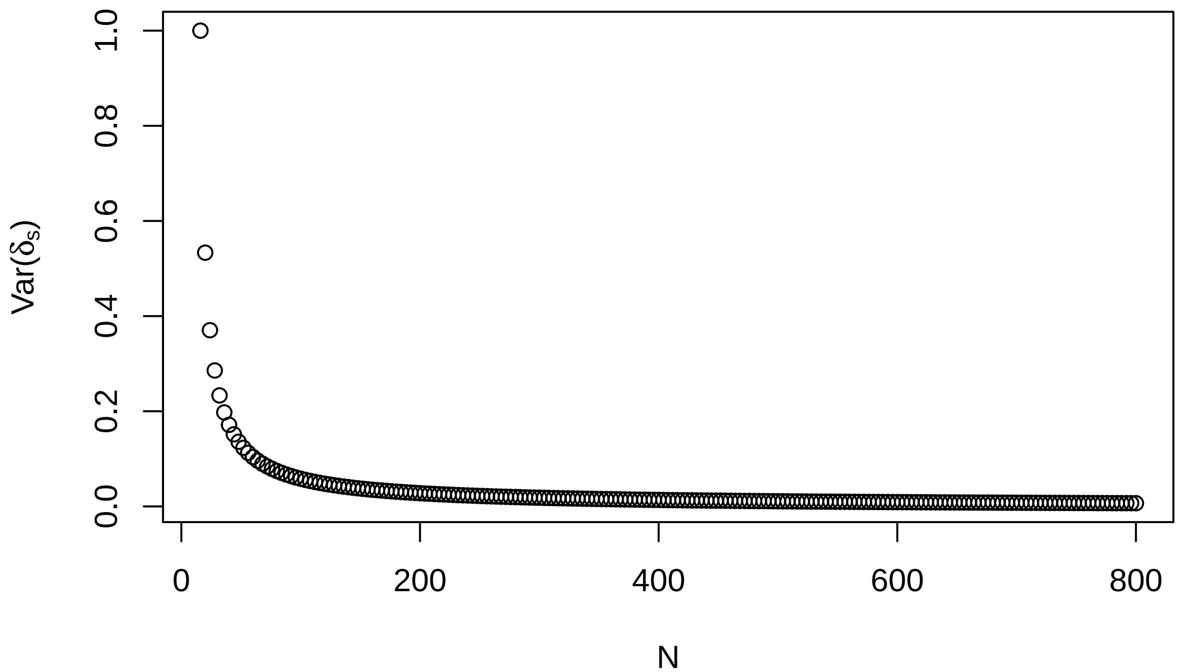


Figure 3. ...

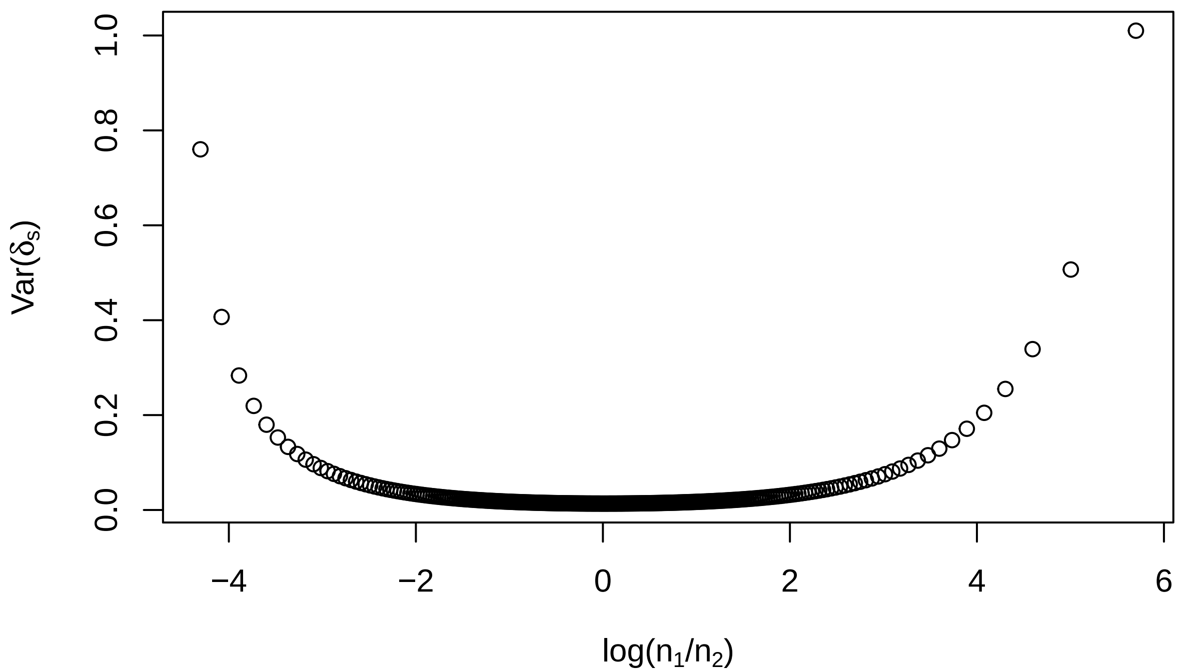


Figure 4. ...

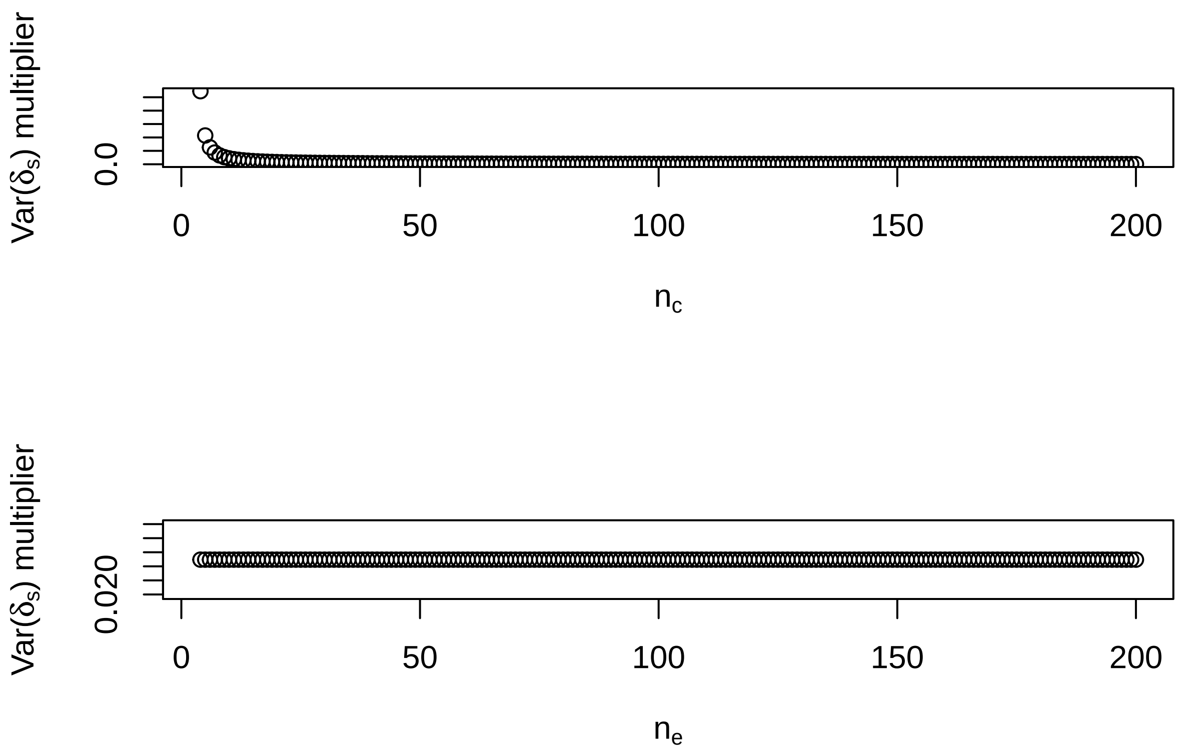

Var(δs) multiplier
0.0
0
50
100
150
200
nc
Var(δs) multiplier
0.020
0
50
100
150
200
ne

Figure 5. ...

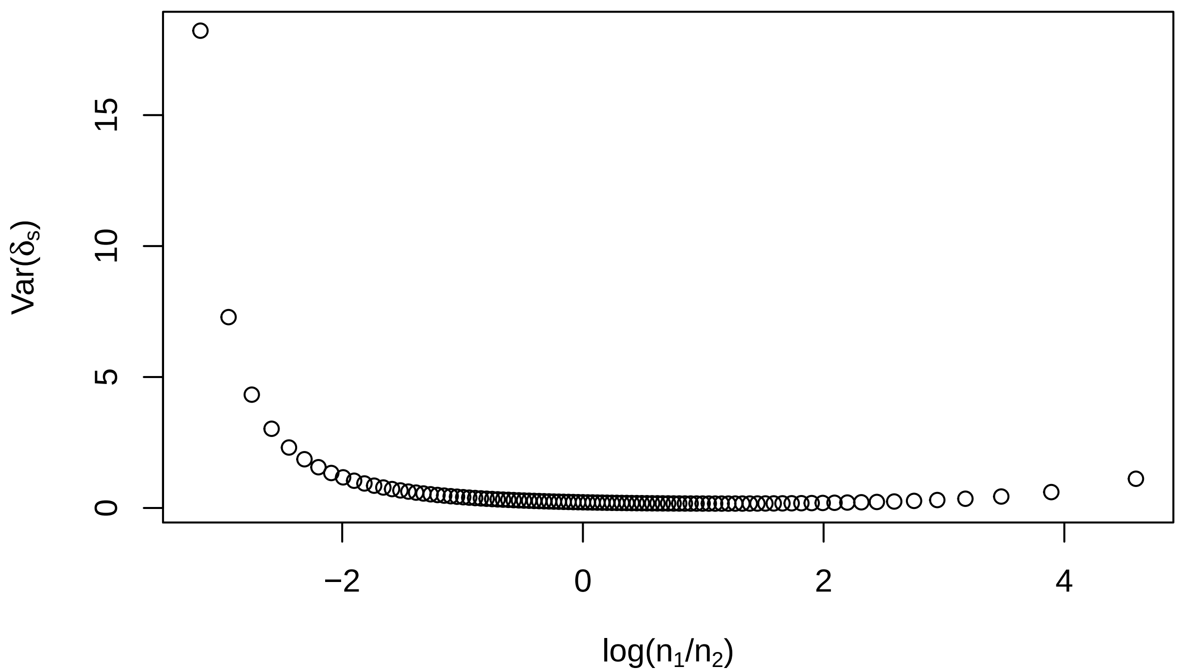


Figure 6. ...

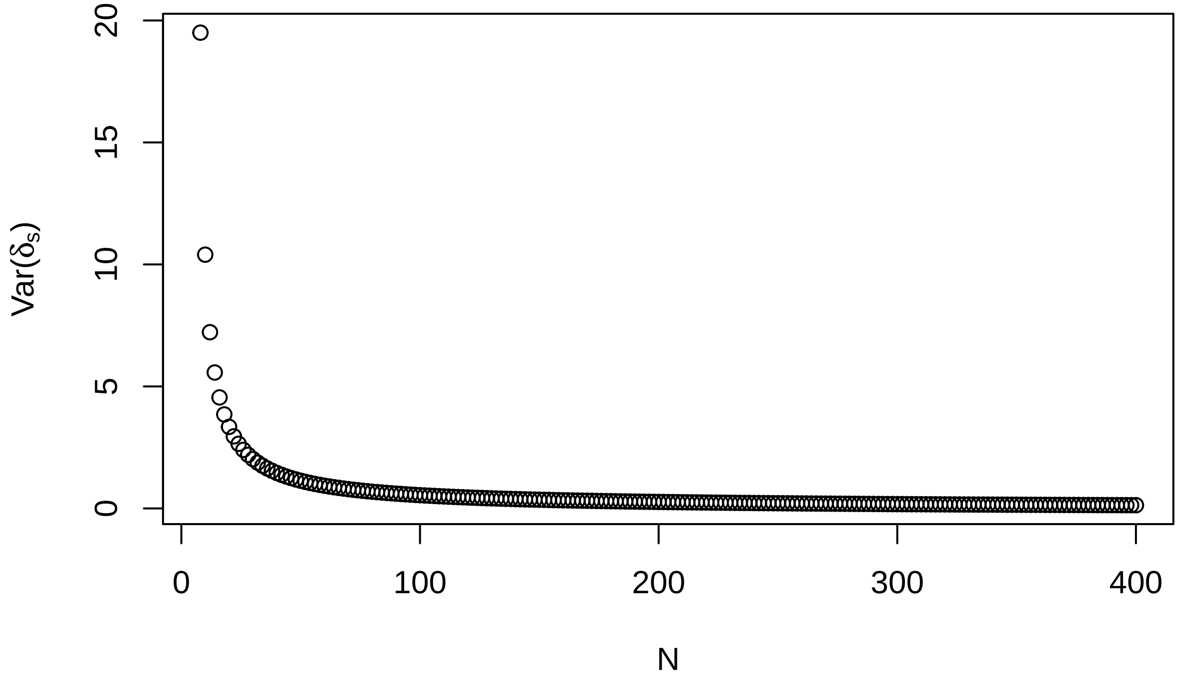


Figure 7. . . .

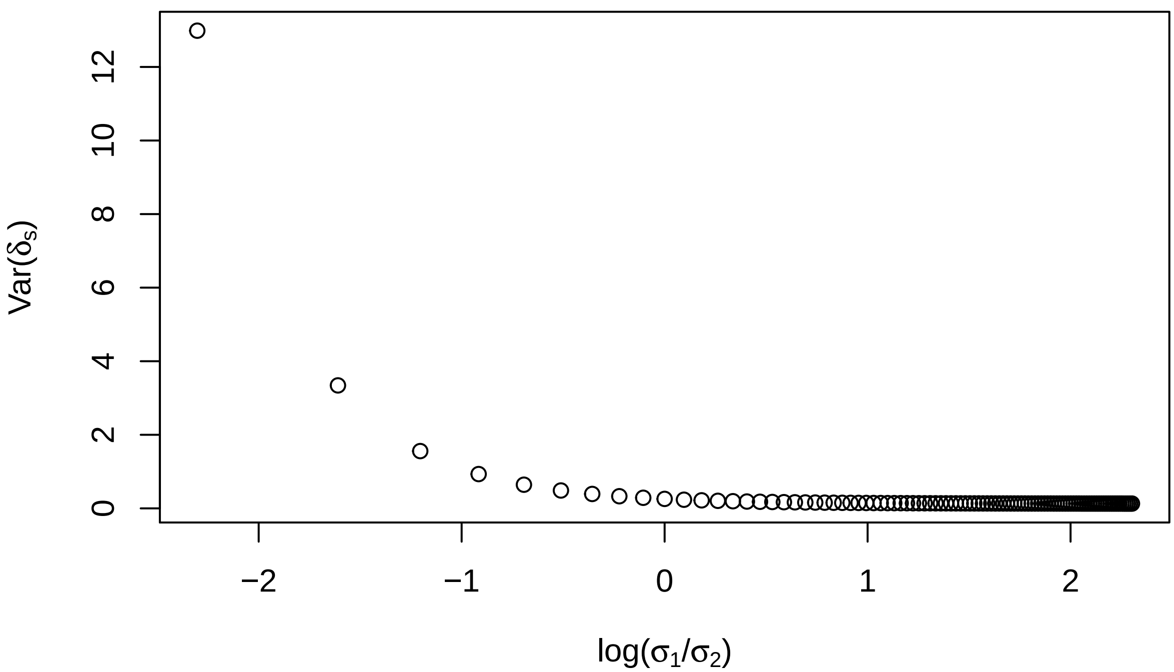


Figure 8. ...

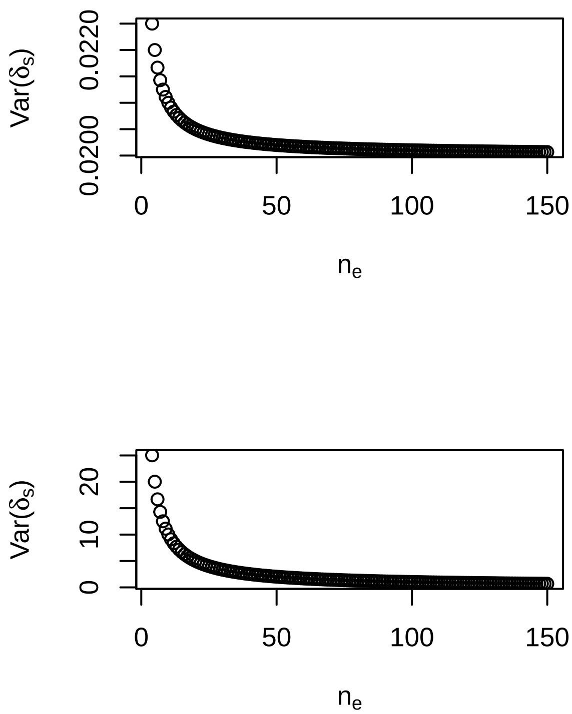
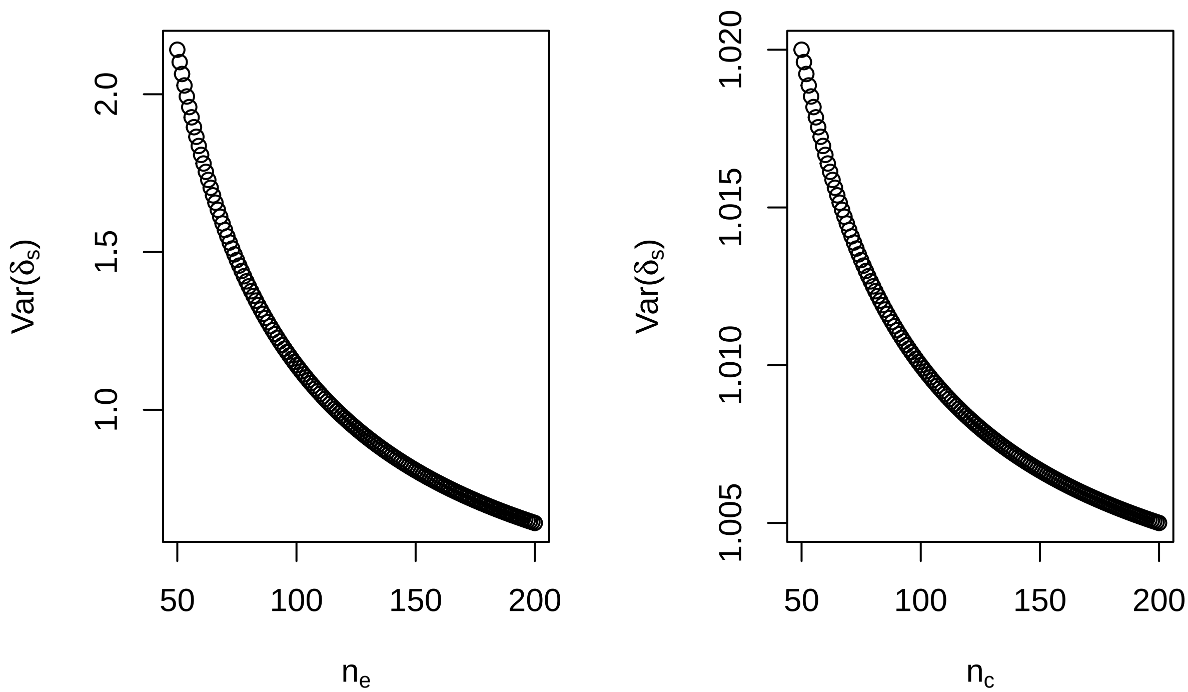

Var(δs)
0.0200
0.0220
0
50
100
150
ne
Var(δs)
0.00
0.15
0
50
100
150
nc
Var(δs)
0
10
20
0
50
100
150
ne
Var(δs)
2.00
2.15
0
50
100
150
nc

Figure 9. . . .



Var(δs)
2.0
1.5
1.0
50
100
150
200
ne
Var(δs)
1.020
1.015
1.010
1.005
50
100
150
200
nc

Figure 10. ...

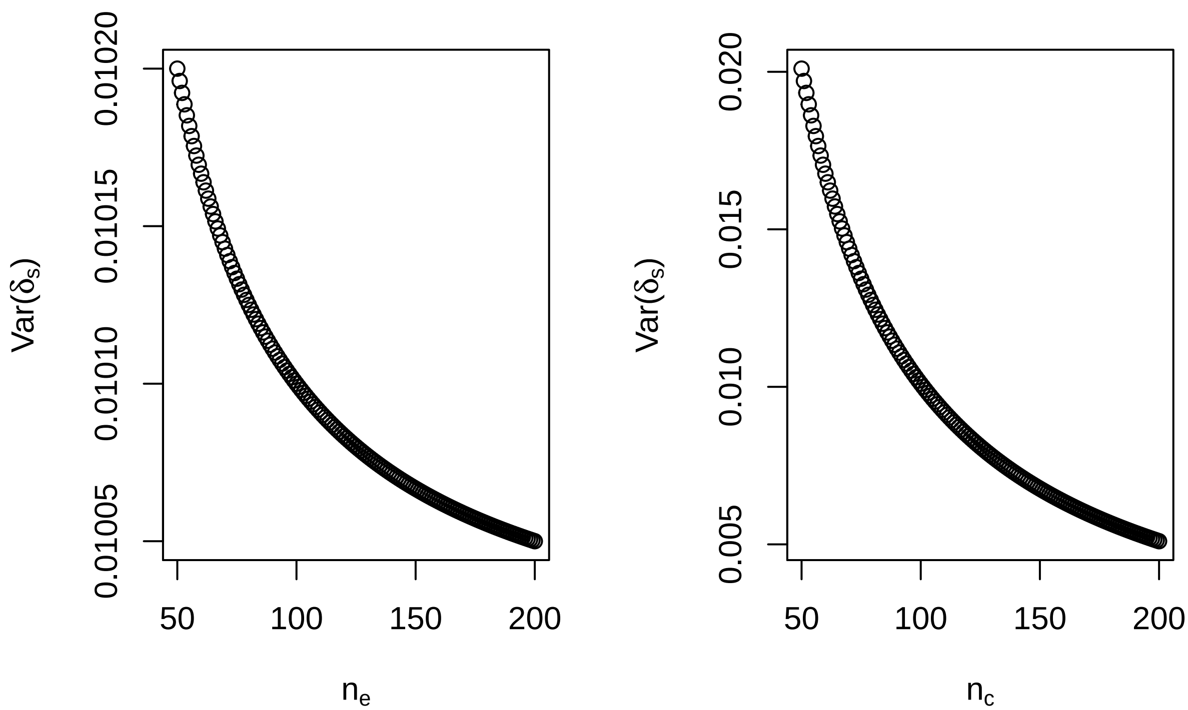

Var(δs)
0.01005
0.01010
0.01015
0.01020
50
100
150
200
ne
Var(δs)
0.005
0.010
0.015
0.020
50
100
150
200
nc

Figure 11. ...

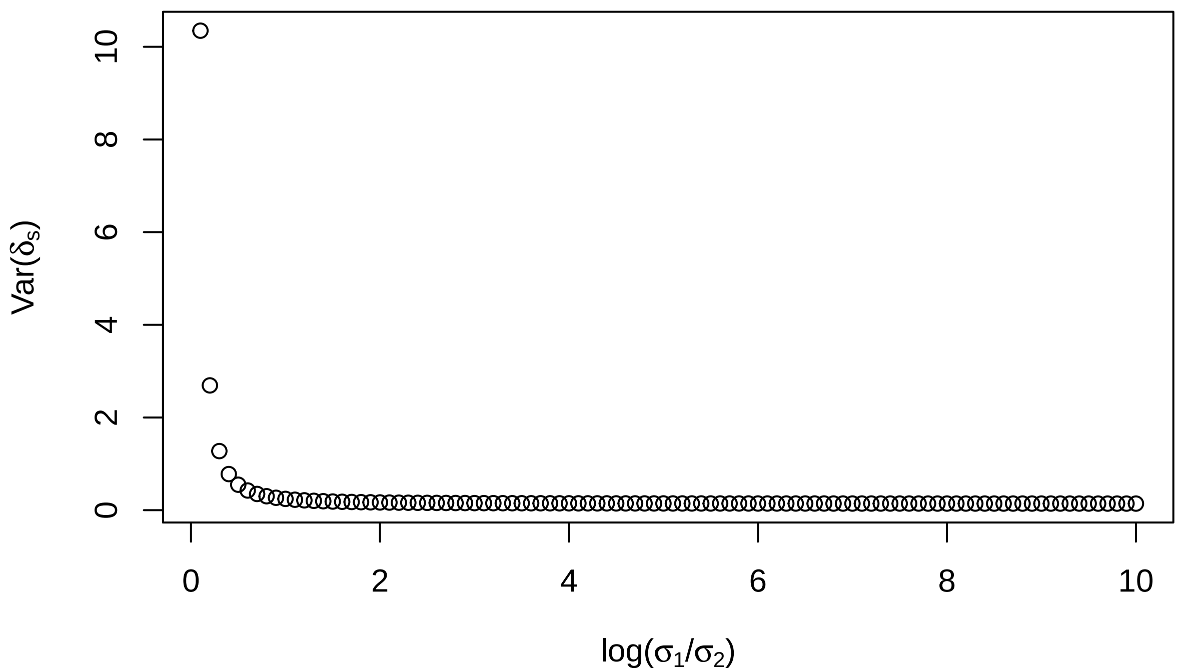


Figure 12. ...