Mathematical study of Glass's d

Marie Delacre[1]

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

Author Note

I would like to thank Matt Williams and Thom Baguley for their helpful insights in order to undertand the phenomenon explained in this appendix.

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

Mathematical study of Glass's d

When two samples are extracted from distributions with identical shapes, with $\sigma_1 = \sigma_2$ and $n_1 = n_2$

When population distributions are symmetric (i.e. $\gamma_1 = 0$), the sampling distribution of glass's d_s is the same, whatever one chooses s_1 or s_2 as standardizer. As an example, in Figure 1, we plotted the sampling distribution of both measures of glass's d_s when two samples of 20 subjects are extracted from two symmetric distributions where $\gamma_1 = 0$,$\gamma_2 = 95.75$, $\sigma_1 = \sigma_2 = 1$ and $\mu_2 = 0$. μ_1 is either 0 or 1, depending on the plot. One can see that in the two plots, distributions of glass's d_S using s_1 and s_2 as standardiser are superimposed.

However, when population distributions are skewed (i.e. $\gamma_1 \neq 0$), the sampling distribution of glass's d_s varies as a function of the chosen standardizer, as illustrated in Figure 2.

It might seem surprising, or even counter-intuitive, as s_1 and s_2 are both estimates of the same population standard deviation (σ), based on the same number of observations (as $n_1 = n_2$), but this phenomenon can be mathematically explained. In the following section, we will provides detailed informations to understand the results plotted in Figure 2.

When distribution is right-skewed, and $\mu_1 - \mu_2 = 0$ (top right plot in Figure 2)

We will first study the configuration where both samples are extracted from a right-skewed distribution where $\mu = 0$, $\sigma = 1$, $\gamma_1 = 6.32$ and $\gamma_2 = 95.75$. Because this distributions is right-skewed, the sampling distributions of $\bar{X}_1$ and $\bar{X}_2$ will also be right-skewed. However, because $\bar{X}_1$ and $\bar{X}_2$ are identically distributed, $\bar{X}_1 - \bar{X}_2$ will follow a symmetric distribution, as illustrated in Figure 3 (right plot). Moreover, it will be centered around $\mu_1 - \mu_2 = 0$, meaning that 50 percent of the mean difference estimates will be positive (i.e. $\bar{X}_1 - \bar{X}_2 > 0$; see green area) and the other 50 percent will be negative

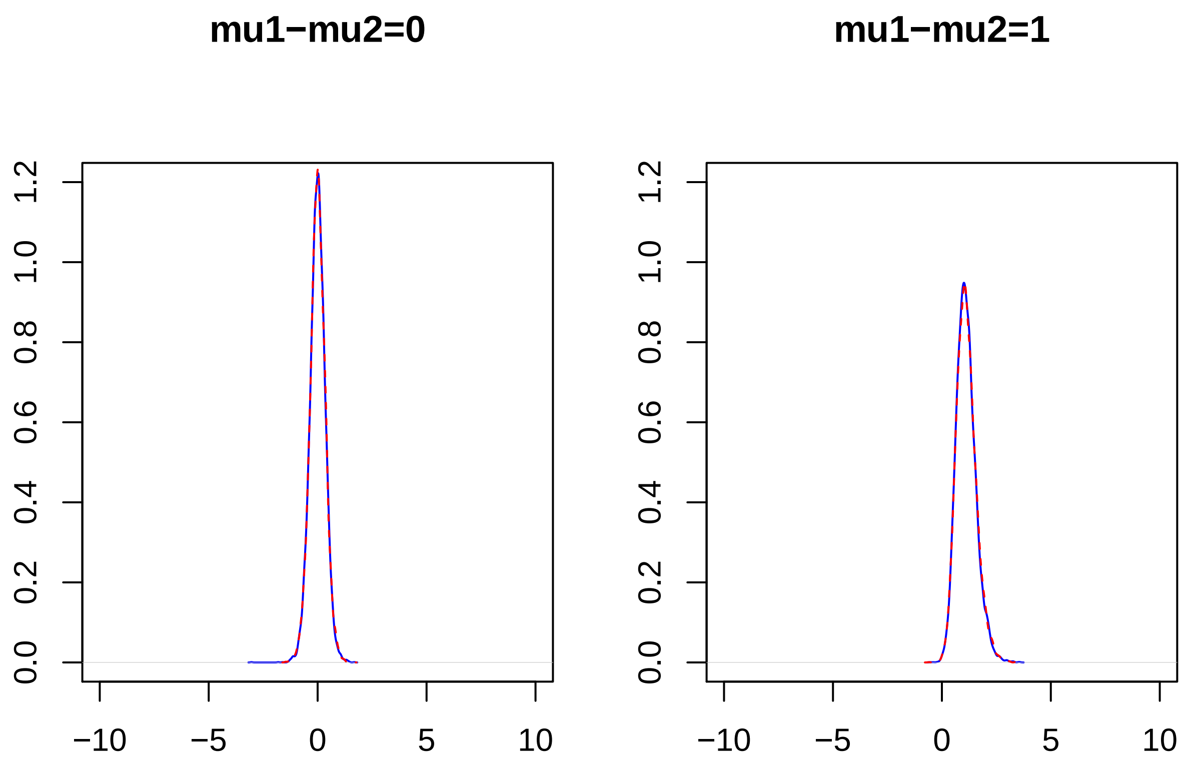

Figure 1. Comparison of Glass's ds when choosing either s1 (blue line) or s2 (red dotted line) as standardizer, with s1=standard deviation of the first sample and s2=standard deviation of the second sample, when n1=n2=20 and both samples are extracted from a distribution where G1 =0, G2=95.75 and sigma=1

(i.e. $\bar{X}_1 - \bar{X}_2 < 0$; see blue area).

Because we compute the mean difference as the mean estimate of the first sample minus the mean estimate of the second sample, there is a positive correlation between $\bar{X}_1$ and $\bar{X}_1 - \bar{X}_2$, and a negative correlation between $\bar{X}_2$ and $\bar{X}_1 - \bar{X}_2$ (correlations would be trivially reversed if we computed $\bar{X}_2 - \bar{X}_1$ instead of $\bar{X}_1 - \bar{X}_2$).

The sampling distributions of s_1 and s_2 are right-skewed, because estimates of the standard deviation are bounded: they can be very large, but never below 0. Moreover, as s_1

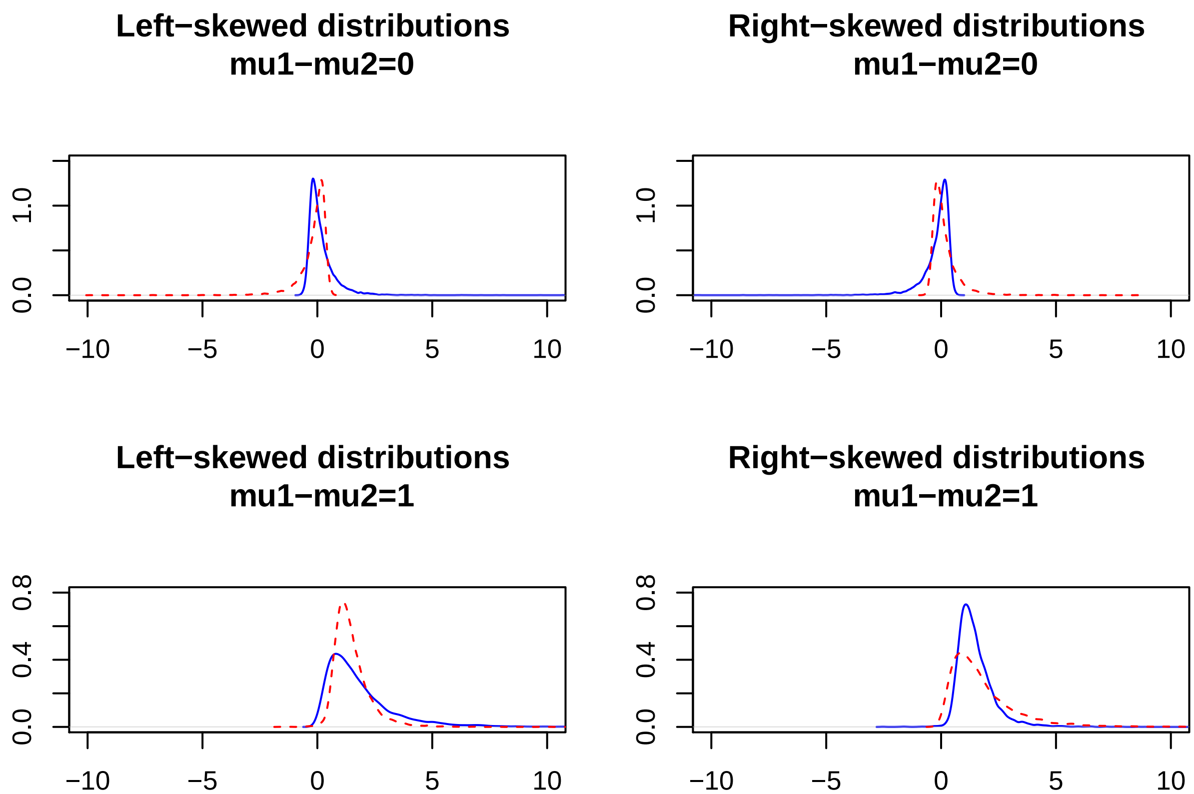

Figure 2. Comparison of Glass's ds when choosing either sd1 (blue line) or sd2 (red dotted line) as standardizer when n1=n2=20 and both samples are extracted from a distribution where sigma=1, G2=95.75, G1 is either -6.32 (left) or 6.32 (right). In all cases, the second sample is extracted from a population distribution where mu2=0. First sample is extracted from a population distribution where mu1 is either 0 (top) of 1 (bottom)

and s_2 are estimates of the same population standard deviation σ, based on the same sample size, of course, the sampling distributions of s_1 and s_2 will be identical, as illustrated in Figure 4.

Therefore, how to explain the different sampling distributions of glass's d_s, as a function of the standardizer? This is due to the fact that when distributions are skewed, there is a non-nul correlation between $\bar{X}$ and s (see Zhang, 2007). More specifically, when distributions are right-skewed, there is a **positive** correlation between $\bar{X}$ and s.

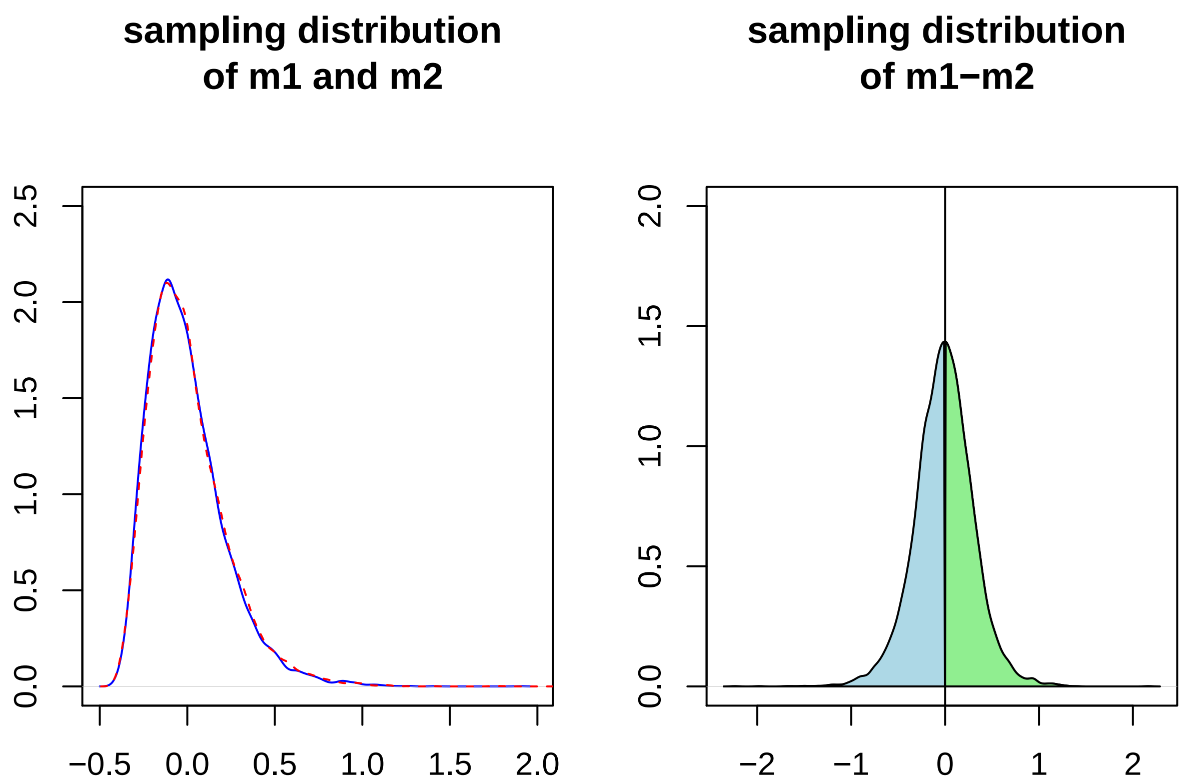

Figure 3. Sampling distribution of m1 (blue line in left plot), m2 (red dotted line in left plot), and m1-m2 (right plot), when m1 and m2 are estimates of the mean of a population distribution where mu=0, sigma=1,G1=6.32 and G2=95.75, with n1=n2=20

First, consider the glass's d_s estimate using s_1 as standardiser. We already mentioned that there is a *positive* correlation between $\bar{X}_1$ and $\bar{X}_1 - \bar{X}_2$ ($cor(\bar{X}_1, \bar{X}_1 - \bar{X}_2) > 0$). Because there is also a positive correlation between $\bar{X}_1$ and s_1 ($cor(\bar{X}_1, s_1) > 0$), it results in a **positive** correlation between $\bar{X}_1 - \bar{X}_2$ and s_1 ($cor(\bar{X}_1 - \bar{X}_2, s_1) > 0$): when moving from the left to the right in the right plot in Figure 3, s_1 get larger. As a consequence, the mean difference estimates in the left tail of the plot (i.e. the most extreme negative estimates) will be divided by a smaller positive value (resulting in a larger ratio) than the mean difference estimates in the right tail of the plot (i.e. the most extreme positive estimates), resulting in a left-skewed sampling distribution of glass's d_S. Importantly, while the median of the

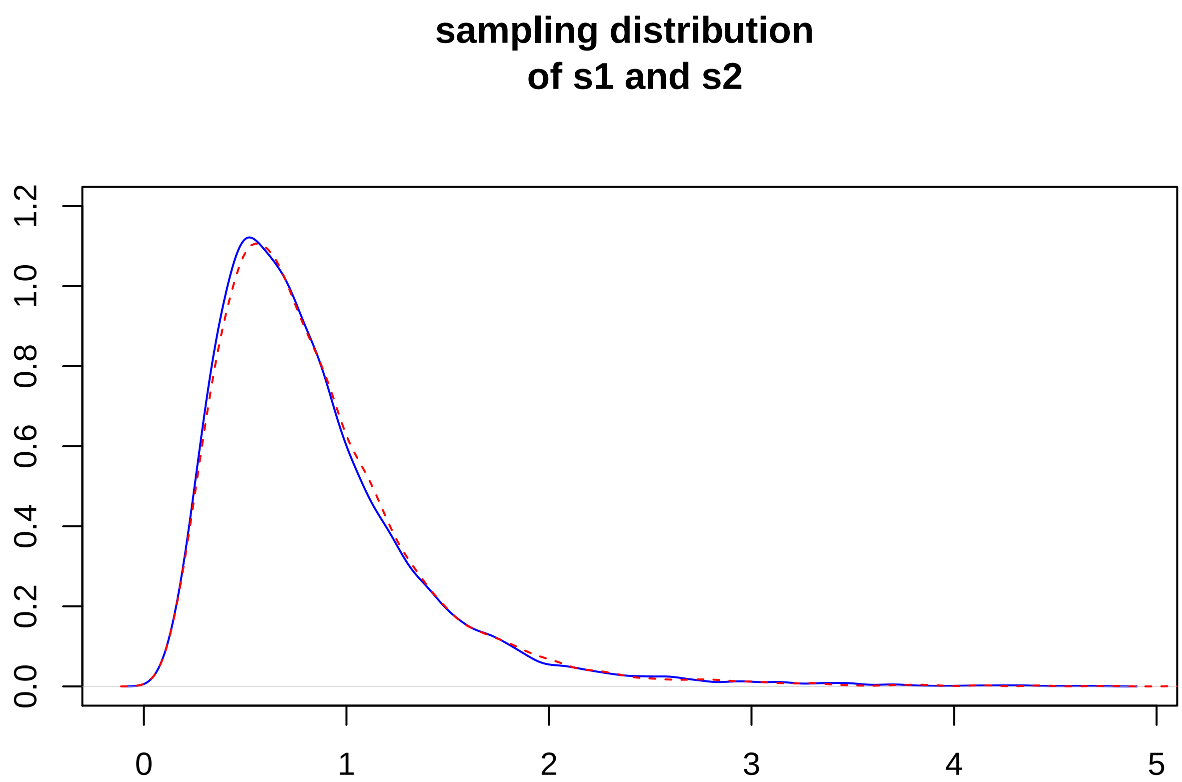

Figure 4. Sampling distribution of s1 (blue line) and s2 (red dotted line), when s1 and s2 are estimates of the standard deviation of a population distribution where mu=0, sigma=1,G1=6.32 and G2=95.75, with n1=n2=20

sampling distribution of glass's d_s is 0, as expected (because the sampling distributions of $\bar{X}_1 - \bar{X}_1$ is centered around 0), the mean will be a little lower (i.e. -0.18), meaning that glass's d_s is negatively biased.

When considering s_2 as standardiser, because there is a *negative* correlation between $\bar{X}_2$ and $\bar{X}_1 - \bar{X}_2$., there is also a **negative** correlation between $\bar{X}_1 - \bar{X}_2$ and s_2: when moving from the left to the right in the right plot in Figure 3, s_2 get lower. In other word, the mean difference estimates in the left tail of the plot will be divided by a larger positive value (resulting in a smaller ratio) than the mean difference estimates in the right tail of the plot, resulting in a right-skewed sampling distribution of glass's d_S. This time, while the median of the sampling distribution of glass's d_s is still 0, the mean will be a little larger (i.e. 0.16), meaning that glass's d_s is positively biased.

When distribution is left-skewed, and $\mu_1 - \mu_2 = 0$ (top left plot in Figure 2)

When distributions are left-skewed, there is a **negative** correlation between $\bar{X}$ and s and therefore, when moving from the left to the right in the right plot in Figure 3, s_1 get lower ($cor(\bar{X_1}, s_1) < 0$ and $cor(\bar{X_1}, \bar{X_1} - \bar{X_2} > 0) \rightarrow cor(\bar{X_1} - \bar{X_2}, s_1) < 0$) and s_2 get larger ($cor(\bar{X_2}, s_2) < 0$ and $cor(\bar{X_2}, \bar{X_1} - \bar{X_2} < 0) \rightarrow cor(\bar{X_1} - \bar{X_2}, s_2) > 0$). As a consequence, when dividing the mean difference by s_1, the estimates of $\mu_1 - \mu_2$ in the left tail of the right plot in Figure 3 (i.e. the most extreme negative estimates) will be divided by a larger positive value (resulting in a smaller ratio) than the ones in the right tail. On the other side, when the mean difference is divided by s_2, the estimates in the left tail of the plot will be divided by a smaller positive value (resulting in a larger ratio) than the ones in the right tail. Unlike what occured when samples were extracted from a right-skewed distribution, when they are extracted from a left-skewed distribution, glass's d_S will be positively biased when using s_1 as a standardiser, and negatively biased when using s_2 as a standardiser.

When distribution is skewed, and $\mu_1 - \mu_2 = 1$ (bottom pslot in Figure 2)

We will first consider the example where both samples are extracted from right-skewed distributions with μ_1 and μ_2 being respectively 1 and 0, and other moments of the population distributions being equal: $\sigma = 1$, $\gamma_1 = 6.32$ and $\gamma_2 = 95.75$ (see bottom right plot in Figure 2).

Of course, the sampling distributions of $\bar{X_1}$ and $\bar{X_2}$ are not superimposed anymore, because $\bar{X_1}$ will be centered around $\mu_1 = 1$, and $\bar{X_2}$ will be centered around $\mu_2 = 0$. However, except for the mean, all other moments of both distributions (i.e. γ_1, γ_2 and σ) remain identical (see left plot in Figure 5) and therefore, the sampling distribution of $\bar{X_1} - \bar{X_2}$ still follow a symmetric distribution, as illustrated in the right plot in Figure 5.

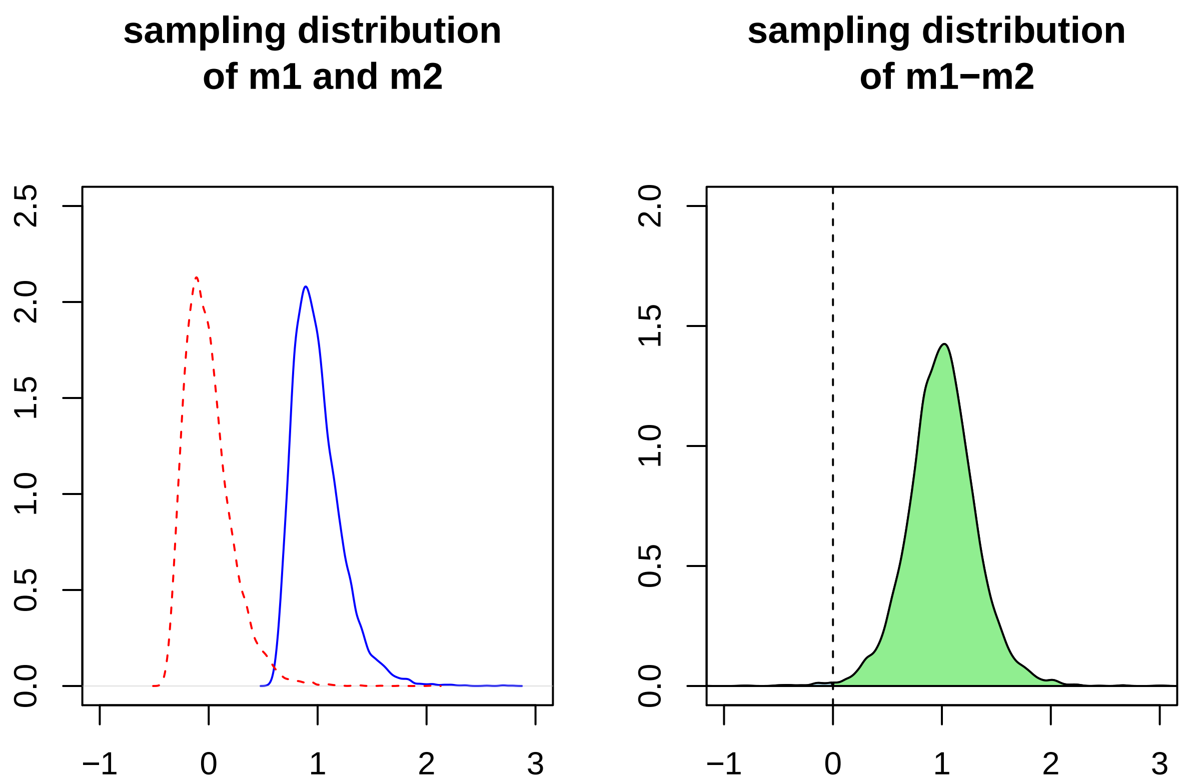

Figure 5. Sampling distribution of m1 (blue line in left plot), m2 (red dotted line in left plot), and m1-m2 (right plot), when m1 is the mean of a sample extracted from a population distribution where mu=1 and m2 is the mean of a sample extracted from a population distribution where mu2=0. Except for the mean, all other moments of both populations distributions are identical, i.e. sigma=1,G1=6.32 and G2=95.75, with n1=n2=20

In previous examples where $\mu_1 - \mu_2$ was nul, because the sampling distribution of $\bar{X}_1 - \bar{X}_2$ was symmetrically centered around 0, the magnitude of the mean difference estimates were the same in both tails. More generally, for a constant k, $|(\mu_1 - \mu_2) - k| = |(\mu_1 - \mu_2) + k|$. Comparing the magnitude of glass's d_s when $\bar{X}_1 - \bar{X}_2 = (\mu_1 - \mu_2) \pm k$ was therefore only a function of the denominator. When $\mu_1 - \mu_2 \neq 0$, comparing the magnitude of glass's d_s when $\bar{X}_1 - \bar{X}_2 = (\mu_1 - \mu_2) \pm k$ is a function of both numerator and denominator.

When $\mu_1 - \mu_2 = 1$, only about 0.39% of the mean estimates are negative, meaning that almost all mean difference estimates will be positive (so will be glass's d_s estimates). When computing glass's d_s using s_1 as standardizer, the mean difference estimates that are close of 0 will be divided by a smaller standard deviation estimate that larger mean difference estimates. On the other side, when computing glass's d_s using s_2 as standardizer, the mean difference estimates that are very small will be divided by a larger standard deviation estimate than large mean difference estimates. It is therefore not surprising that the variance of the sampling distribution of glass's d_s is larger when using s_2 rather than s_1 as standardizer. When distributions are extracted from a left-skewed distribution (bettom left in Figure 2), this is exactly the opposite.

When two samplesare extracted from distributions with identical shapes, and $n_1 \neq n_2$