

¹ Mathematical study of Glass's d

² Marie Delacre¹

³ ¹ Université Libre de Bruxelles, Service of Analysis of the Data (SAD), Bruxelles, Belgium

⁴ Author Note

⁵ I would like to thank Matt Williams and Thom Baguley for their helpful insights in
⁶ order to understand the phenomenon explained in this appendix.

⁷ Correspondence concerning this article should be addressed to Marie Delacre, CP191,
⁸ avenue F.D. Roosevelt 50, 1050 Bruxelles. E-mail: marie.delacre@ulb.ac.be

9

Abstract

10

11 *Keywords:* keywords

12 Word count: X

¹³ Mathematical study of Glass's d

¹⁴ **When two samples are extracted from distributions with identical shapes, with**

¹⁵ $\sigma_1 = \sigma_2$ and $n_1 = n_2$

¹⁶ When population distributions are symmetric (i.e. $\gamma_1 = 0$), the sampling distribution of
¹⁷ glass's d_s is the same, whatever one chooses s_1 or s_2 as standardizer. As an example, in
¹⁸ Figure 1, we plotted the sampling distribution of both measures of glass's d_s when two
¹⁹ samples of 20 subjects are extracted from two symmetric distributions where
²⁰ $\gamma_1 = 0, \gamma_2 = 95.75$, $\sigma_1 = \sigma_2 = 1$ and $\mu_2 = 0$. μ_1 is either 0 or 1, depending on the plot. One
²¹ can see that in the two plots, distributions of glass's d_s using s_1 and s_2 as standardiser are
²² superimposed.¹

²³ However, when population distributions are skewed (i.e. $\gamma_1 \neq 0$), the sampling
²⁴ distribution of glass's d_s varies as a function of the chosen standardizer, as illustrated in
²⁵ Figure 2.

²⁶ It might seem surprising, or even counter-intuitive, as in all plots, s_1 and s_2 are both
²⁷ estimates of the same population standard deviation (σ), based on the same number of
²⁸ observations (as $n_1 = n_2$), but this phenomenon can be mathematically explained. In the
²⁹ following section, we will provides detailed informations to understand the results plotted in
³⁰ Figure 2.

¹ Anytime the mean difference is null, the sampling distribution of glass will be symmetric. On the other side, anytime the mean difference is non null, the sampling distribution of glass's ds will be skewed: right-skewed if the mean difference is positive, left-skewed if the mean difference is negative. The right skewed distribution is therefore only due to the sense of the mean difference

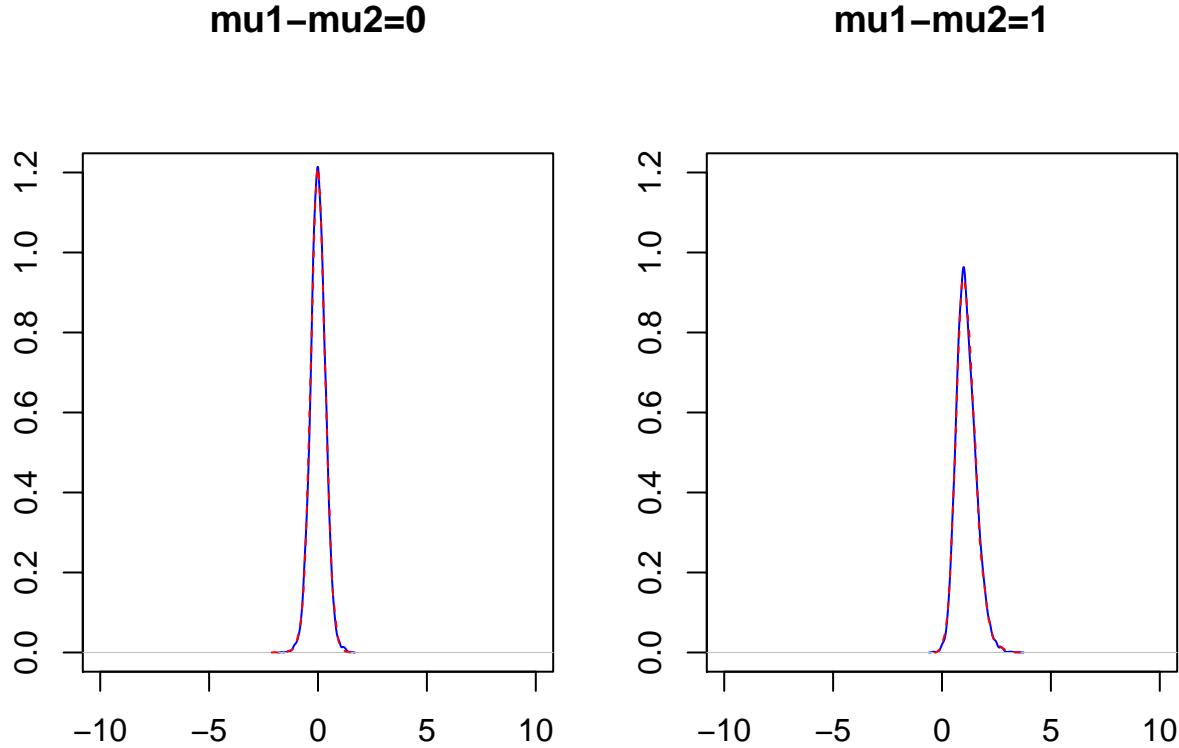


Figure 1. Comparison of Glass's ds when choosing either s_1 (blue line) or s_2 (red dotted line) as standardizer, with s_1 =standard deviation of the first sample and s_2 =standard deviation of the second sample, when $n_1=n_2=20$ and both samples are extracted from a distribution where $G_1 = 0$, $G_2=95.75$ and $\sigma=1$

- ³¹ When both samples are extracted from a common right-skewed distribution
- ³² $(\mu_1 - \mu_2 = 0)$ (top right plot in Figure 2)

³³ We will first study the configuration where both samples are extracted from a
³⁴ right-skewed distribution where $\mu = 0$, $\sigma = 1$, $\gamma_1 = 6.32$ and $\gamma_2 = 95.75$. Because this
³⁵ distribution is right-skewed, the sampling distributions of \bar{X}_1 and \bar{X}_2 will also be
³⁶ right-skewed. However, because \bar{X}_1 and \bar{X}_2 are identically distributed, $\bar{X}_1 - \bar{X}_2$ will follow a
³⁷ symmetric distribution, as illustrated in Figure 3 (right plot). More specifically, the
³⁸ distribution will be symmetrically centered around $\mu_1 - \mu_2 = 0$ (i.e. the green area in the

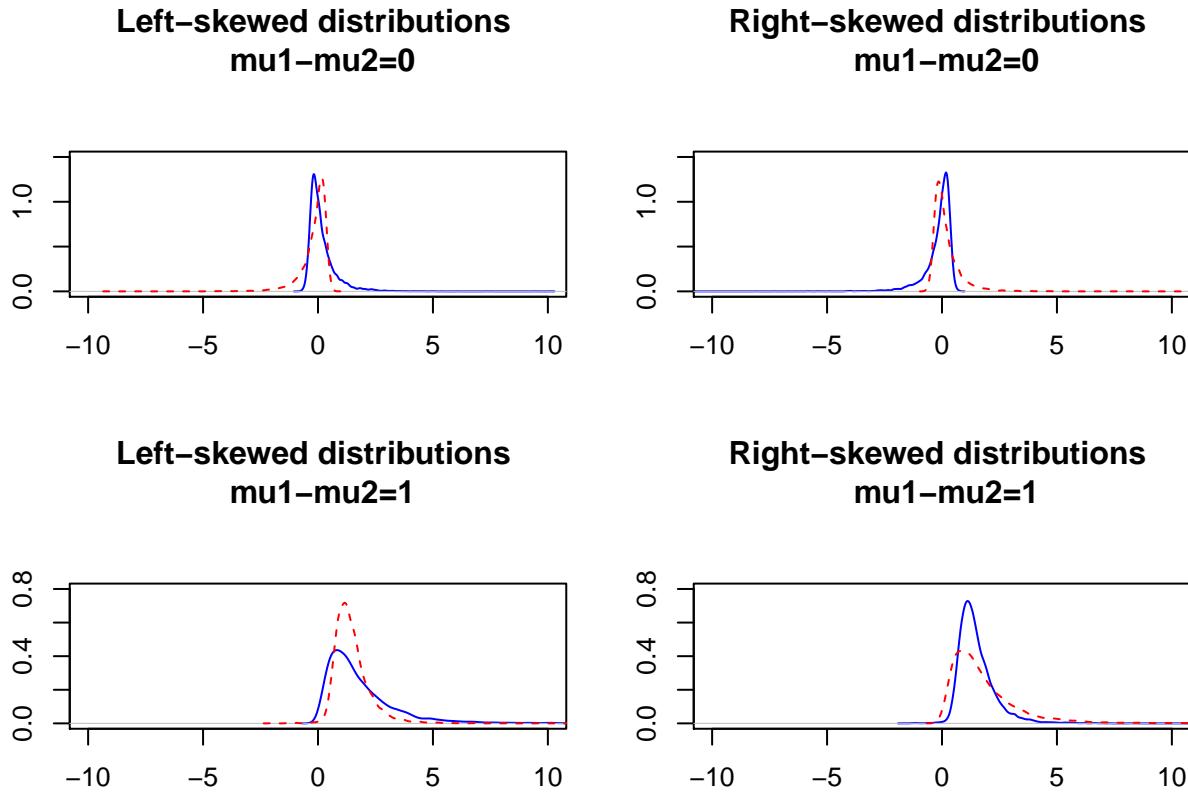


Figure 2. Comparison of Glass's ds when choosing either sd1 (blue line) or sd2 (red dotted line) as standardizer when $n_1=n_2=20$ and both samples are extracted from a distribution where $\sigma=1$, $G_2=95.75$, G_1 is either -6.32 (left) or 6.32 (right). In all cases, the second sample is extracted from a population distribution where $\mu_2=0$. First sample is extracted from a population distribution where μ_1 is either 0 (top) or 1 (bottom)

39 right plot in Figure 3 is the reflexion of the blue area, with the vertical line being the line of
40 reflexion). It means that:

- 41 - Half the mean difference estimates will be positive (i.e. $\bar{X}_1 - \bar{X}_2 > 0$; see green area)
42 and the other half will be negative (i.e. $\bar{X}_1 - \bar{X}_2 < 0$; see blue area).
43 - For a constant k , $|(\mu_1 - \mu_2) - k| = |(\mu_1 - \mu_2) + k|$

44 Because we compute the mean difference as the mean estimate of the first sample
45 minus the mean estimate of the second sample, there is a positive correlation between \bar{X}_1

⁴⁶ 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
⁴⁷ trivially reversed if we computed $\bar{X}_2 - \bar{X}_1$ instead of $\bar{X}_1 - \bar{X}_2$).

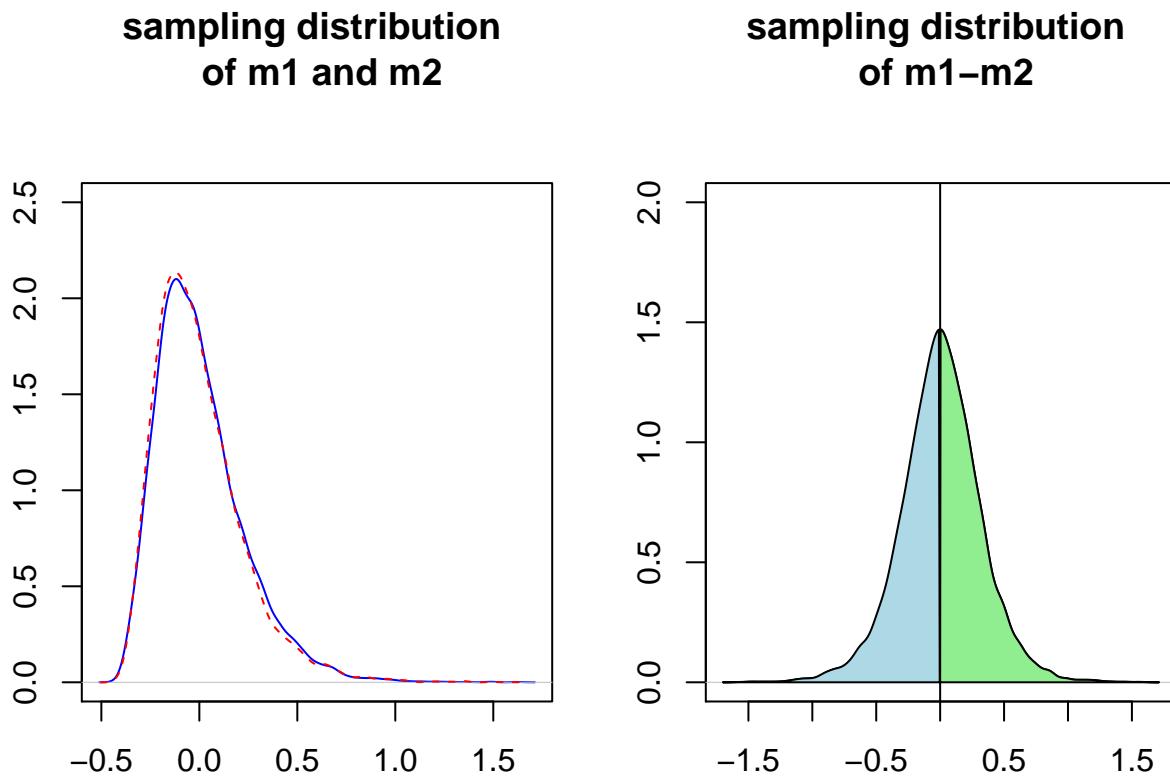


Figure 3. Sampling distribution of m_1 (blue line in left plot), m_2 (red dotted line in left plot), and m_1-m_2 (right plot), when m_1 and m_2 are estimates of the mean of a population distribution where $\mu=0$, $\sigma=1$, $G_1=6.32$ and $G_2=95.75$, with $n_1=n_2=20$

⁴⁸ The sampling distributions of s_1 and s_2 are right-skewed, because estimates of the
⁴⁹ standard deviation are bounded: they can be very large, but never below 0. Moreover, as s_1
⁵⁰ and s_2 are estimates of the same population standard deviation σ , based on the same sample
⁵¹ size, of course, the sampling distributions of s_1 and s_2 will be identical, as illustrated in
⁵² Figure 4.

⁵³ Therefore, how to explain the different sampling distributions of glass's d_s , as a
⁵⁴ function of the standardizer? This is due to the fact that when distributions are skewed,

sampling distribution of s1 and s2

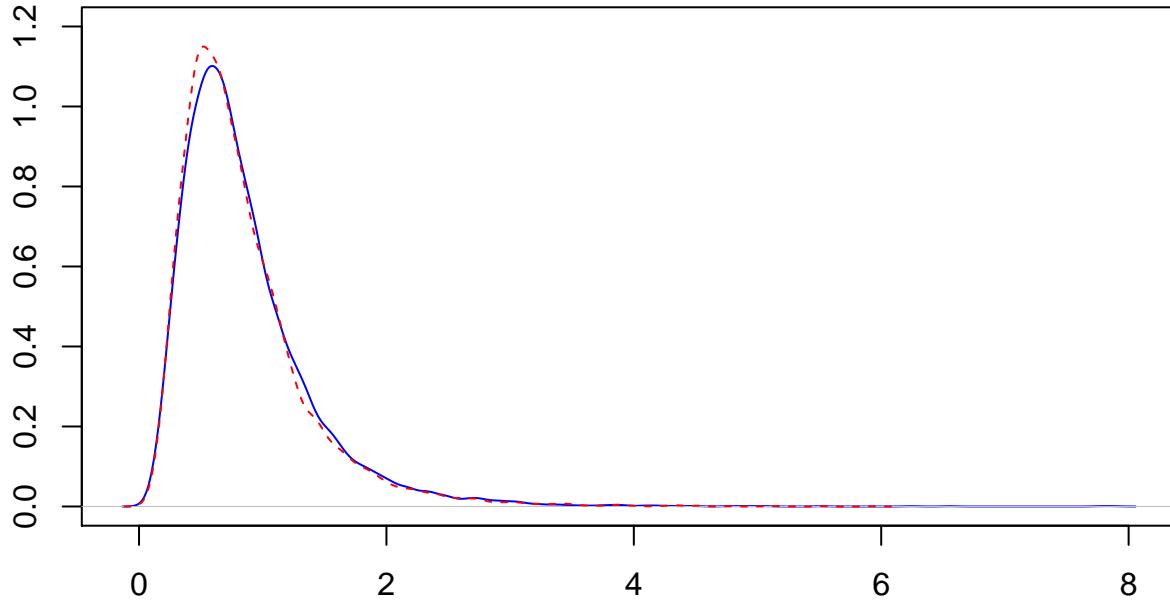


Figure 4. Sampling distribution of s1 (blue line) and s2 (red dotted line), when s1 and s2 are estimates of the standard deviation of a population distribution where mu=0, sigma=1, G1=6.32 and G2=95.75, with n1=n2=20

55 there is a non-nul correlation between \bar{X} and s (see Zhang, 2007). More specifically, when
 56 distributions are right-skewed, there is a **positive** correlation between \bar{X} and s, as
 57 illustrated in the left plots in Figure 5.

58 First, consider the glass's d_s estimate using s_1 as standardiser. We already mentioned
 59 that there is a *positive* correlation between \bar{X}_1 and $\bar{X}_1 - \bar{X}_2$ ($\text{cor}(\bar{X}_1, \bar{X}_1 - \bar{X}_2) > 0$).
 60 Because there is also a positive correlation between \bar{X}_1 and s_1 ($\text{cor}(\bar{X}_1, s_1) > 0$), it results in
 61 a **positive** correlation between $\bar{X}_1 - \bar{X}_2$ and s_1 ($\text{cor}(\bar{X}_1 - \bar{X}_2, s_1) > 0$; see top right plot in
 62 Figure 5): when moving from the left to the right in the right plot in Figure 3, s_1 get larger.

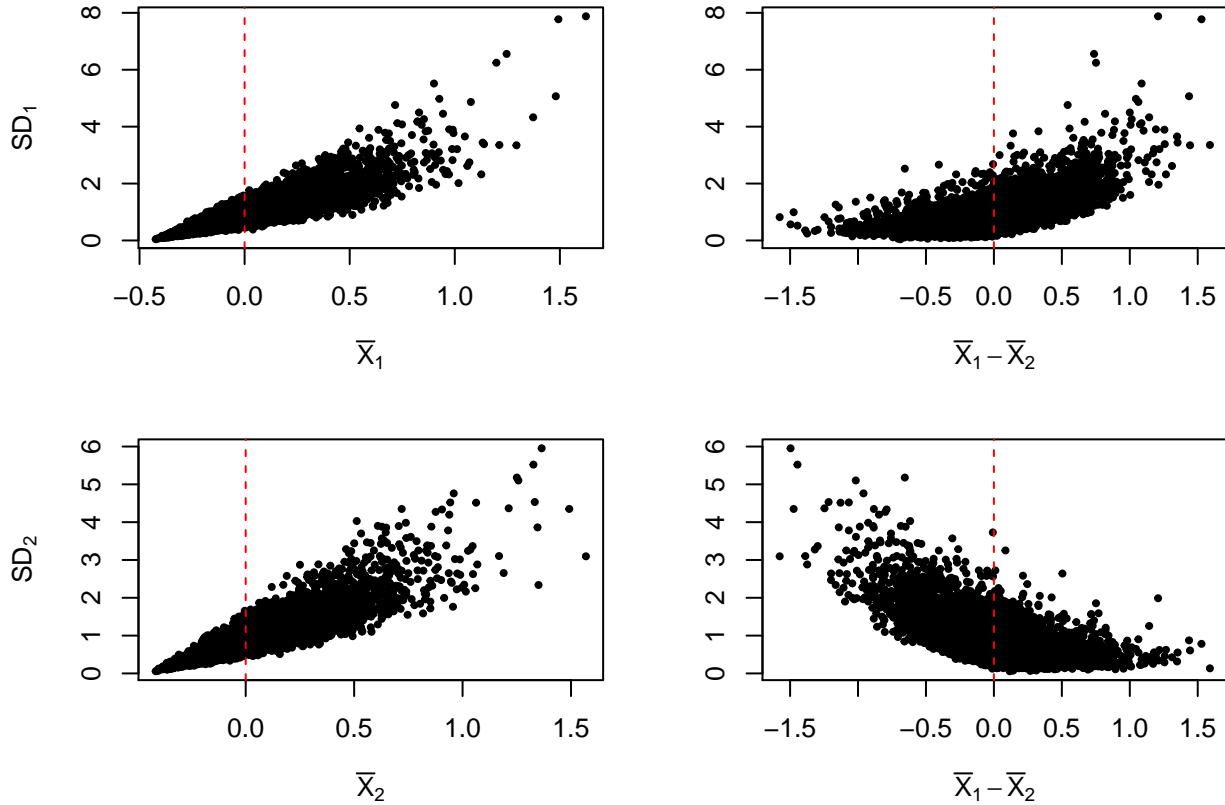


Figure 5. sample mean against sample standard deviation (left) and sample mean difference against sample standard deviation (right), for 100,000 iterations of sampling from two population distributions where $\mu=0$, $\sigma=1$, $G_1=6.32$ and $G_2=95.75$, with $n_1=n_2=20$

63 As a consequence, for many positive numbers k , $-k$ will be divided by a smaller
 64 positive value (resulting in a larger ratio) than $+k$, resulting in a left-skewed sampling
 65 distribution of glass's d_s .² Importantly, while the median of the sampling distribution of
 66 glass's d_s is 0, as expected (because the sampling distributions of $\bar{X}_1 - \bar{X}_1$ is centered around
 67 0), the mean will be a little lower (i.e. -0.16), meaning that glass's d_s is negatively biased.

68 When considering s_2 as standardiser, because there is a *negative* correlation between \bar{X}_2

² This is not true for each value of k , because the correlation between the mean difference and standard deviation estimates is not perfect, but the larger k (i.e. the more extreme the mean difference estimate), the larger the probability that it is true.

69 and $\bar{X}_1 - \bar{X}_2$, there is also a **negative** correlation between $\bar{X}_1 - \bar{X}_2$ and s_2 (see bottom right
 70 plot in Figure 5): when moving from the left to the right in the right plot in Figure 3, s_2 get
 71 lower. In other word, for many positive numbers k , $-k$ will be divided by a larger positive
 72 value (resulting in a larger ratio) than $+k$, resulting in a right-skewed sampling distribution
 73 of glass's d_S . This time, while the median of the sampling distribution of glass's d_s is still 0,
 74 the mean will be a little larger (i.e. 0.18), meaning that glass's d_s is positively biased.

75 **When both samples are extracted from a common left-skewed distribution**

76 $(\mu_1 - \mu_2 = 0)$ (**top left plot in Figure 2**)

77 When distributions are left-skewed, one observes the opposite: there is a **negative**
 78 correlation between \bar{X} and s , as illustrated in the left plots in Figure 6.

79 and therefore, when moving from the left to the right in the right plot in Figure 3, s_1
 80 get lower ($\text{cor}(\bar{X}_1, s_1) < 0$ and $\text{cor}(\bar{X}_1, \bar{X}_1 - \bar{X}_2) > 0 \rightarrow \text{cor}(\bar{X}_1 - \bar{X}_2, s_1) < 0$) and s_2 get
 81 larger ($\text{cor}(\bar{X}_2, s_2) < 0$ and $\text{cor}(\bar{X}_2, \bar{X}_1 - \bar{X}_2) < 0 \rightarrow \text{cor}(\bar{X}_1 - \bar{X}_2, s_2) > 0$). As a
 82 consequence, glass's d_S will be positively biased when using s_1 as a standardiser, and
 83 negatively biased when using s_2 as a standardiser.

84 **When samples are extracted from skewed distributions, with $\mu_1 - \mu_2 = 1$**

85 (**bottom plot in Figure 2**)

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

90 Of course, the sampling distributions of \bar{X}_1 and \bar{X}_2 are not superimposed anymore,
 91 because \bar{X}_1 will be centered around $\mu_1 = 1$, and \bar{X}_2 will be centered around $\mu_2 = 0$.
 92 However, except for the mean, all other moments of both distributions (i.e. γ_1 , γ_2 and σ)

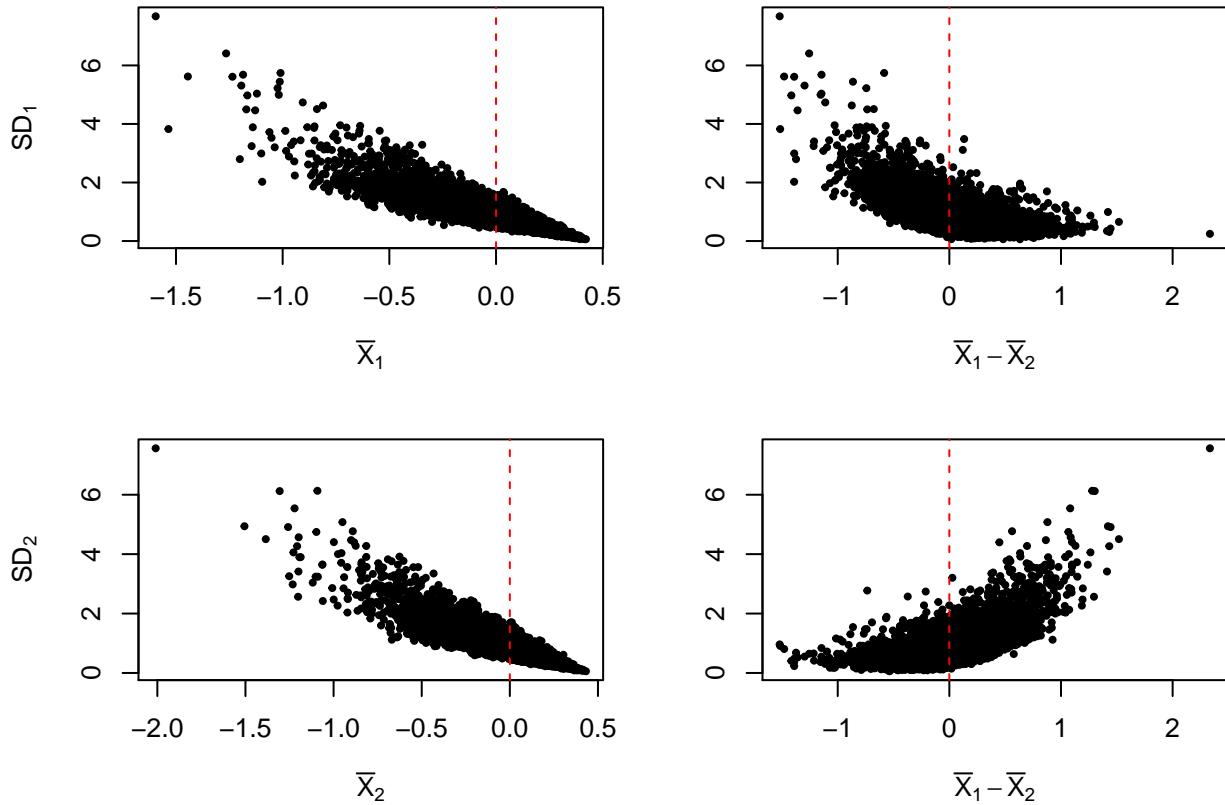


Figure 6. sample mean against sample standard deviation (left) and sample mean difference against sample standard deviation (right), for 100,000 iterations of sampling from two population distributions where $\mu=0$, $\sigma=1$, $G_1=-6.32$ and $G_2=95.75$, with $n_1=n_2=20$

93 remain identical (see left plot in Figure 7) and therefore, the sampling distribution of
 94 $\bar{X}_1 - \bar{X}_2$ still follow a symmetric distribution, as illustrated in the right plot in Figure 7.

95 When $\mu_1 - \mu_2$ was nul, comparing the magnitude of glass's d_s when
 96 $\bar{X}_1 - \bar{X}_2 = (\mu_1 - \mu_2) \pm k$ was only a function of the denominator (as
 97 $|(\mu_1 - \mu_2) - k| = |(\mu_1 - \mu_2) + k|$). When $\mu_1 - \mu_2 \neq 0$, it is a function of both numerator and
 98 denominator. For example, when $\mu_1 - \mu_2 = 1$, only about 0.39% of the mean estimates are
 99 negative, meaning that almost all mean difference estimates will be positive (so will be
 100 glass's d_s estimates). When computing glass's d_s using s_1 as standardizer, the mean
 101 difference estimates that are close of 0 will be divided by a smaller standard deviation

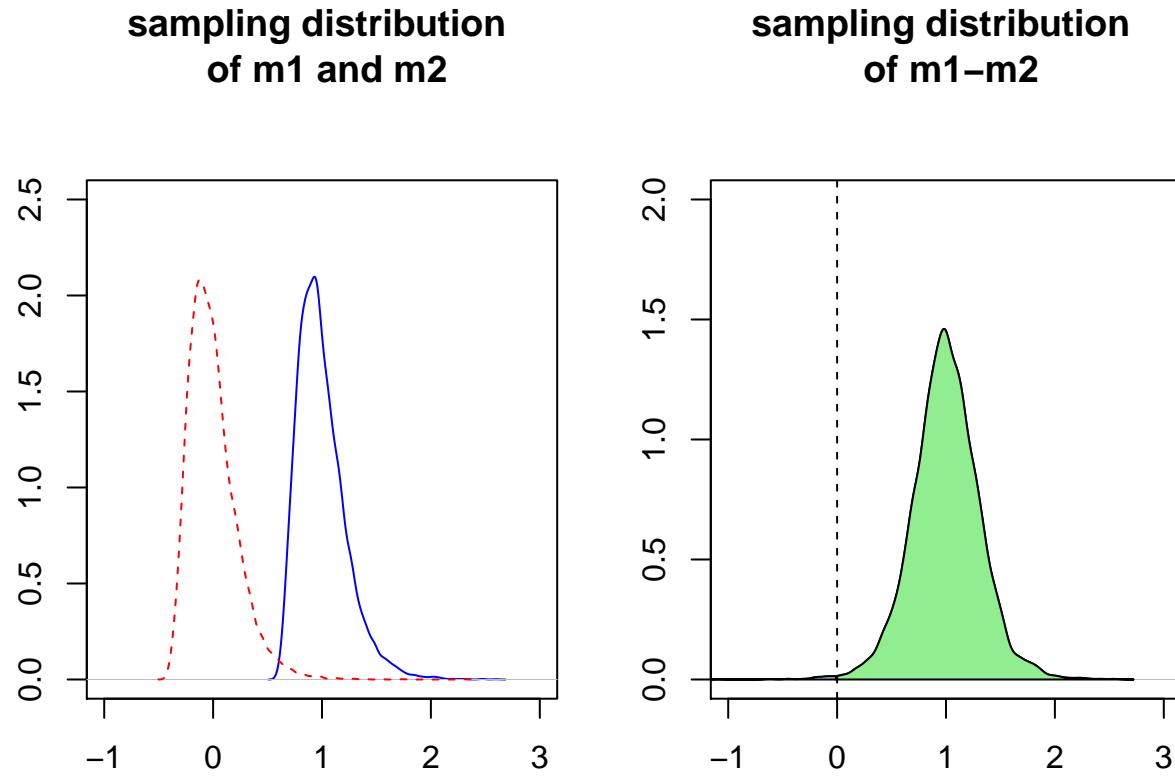


Figure 7. Sampling distribution of m_1 (blue line in left plot), m_2 (red dotted line in left plot), and m_1-m_2 (right plot), when m_1 is the mean of a sample extracted from a population distribution where $\mu=1$ and m_2 is the mean of a sample extracted from a population distribution where $\mu_2=0$. Except for the mean, all other moments of both populations distributions are identical, i.e. $\sigma=1$, $G_1=6.32$ and $G_2=95.75$, with $n_1=n_2=20$

102 estimate that larger mean difference estimates (as $\text{cor}(\bar{X}_1 - \bar{X}_2, s_1) > 0$). On the other side,
 103 when computing glass's d_s using s_2 as standardizer, the mean difference estimates that are
 104 very small will be divided by a larger standard deviation estimate than large mean difference
 105 estimates (as $\text{cor}(\bar{X}_1 - \bar{X}_2, s_2) < 0$). It is therefore not surprising that the sampling
 106 distribution of glass's d_s is more skewed and variable when using s_2 rather than s_1 as
 107 standardizer.³. When distributions are extracted from a left-skewed distribution (bottom left

³ Of course, when the mean difference is negative, this is the opposite: the sampling distribution of glass's d_s

¹⁰⁸ in Figure 2), this is exactly the opposite.

¹⁰⁹ **When two samples are extracted from distributions with identical shapes, and**

¹¹⁰ $n_1 \neq n_2$

¹¹¹ When population distributions are symmetric (i.e. $\gamma_1 = 0$), the sampling distribution of
¹¹² glass's d_s is only a function of the sample size of the group from which standardizer is
¹¹³ computed (because $\sigma_1 = \sigma_2$). Using the SD of the smallest group as standardiser results in a
¹¹⁴ more biased and variable measure of Glass's d_S , whatever it is the first or the second group,
¹¹⁵ as illustrated in Figure 8.

¹¹⁶ We know that the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is always symmetric, but that the
¹¹⁷ sampling distribution of any standard deviation is right-skewed (because a standard
¹¹⁸ deviation can never be below 0). The lower the sample size, the more skewed the sampling
¹¹⁹ distribution of the standard deviation and the larger it's variance. As a consequence, the
¹²⁰ more skewed the sampling distribution of glass's d_s).

¹²¹ When distributions are skewed, the skewness of the glass's d_s distribution (and
¹²² therefore, the bias and variance of glass's d_s) depends on two parameters: the size of the
¹²³ group from which standardizer is computed (as with symmetric distributions), but also the
¹²⁴ correlation between $\bar{X}_1 - \bar{X}_2$ and s_j . As long as $\mu_1 - \mu_2$ is positive (or negative), the more
¹²⁵ skewed and variable estimation of glass's d_s will occur when s_j is simultaneously negatively
¹²⁶ (or positively) correlated with $\bar{X}_1 - \bar{X}_2$ and computed based on the smallest sample size. On
¹²⁷ the other side, the best estimation (less biased and variable) will occur when s_j is
¹²⁸ simultaneously positively (or negatively) correlated with $\bar{X}_1 - \bar{X}_2$ and computed based on
¹²⁹ the largest sample size. This is illustrated in Figure ?? in four plots where samples are
¹³⁰ extracted from either a right-skewed distribution ($\gamma_1 = 6.32$; right) or a left-skewed
¹³¹ distribution ($\gamma_1 = -6.32$; left), and with a positive (top) or negative (bottom) population

will be left-skewed, and will be more skewed and variable when using s_1 rather than s_2 as standardizer

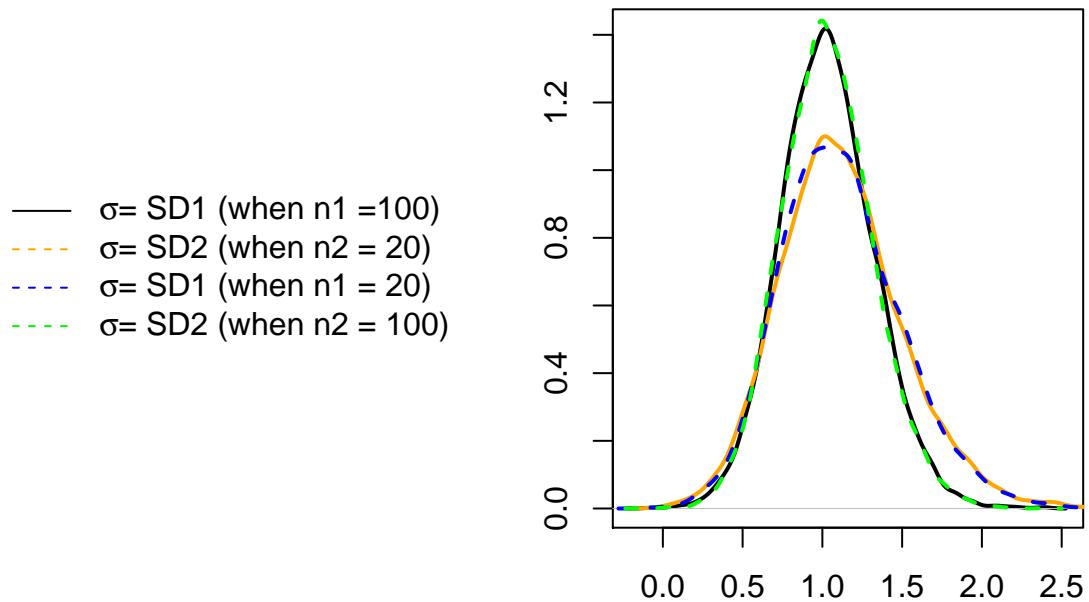
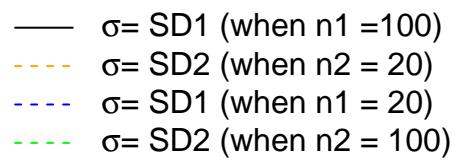
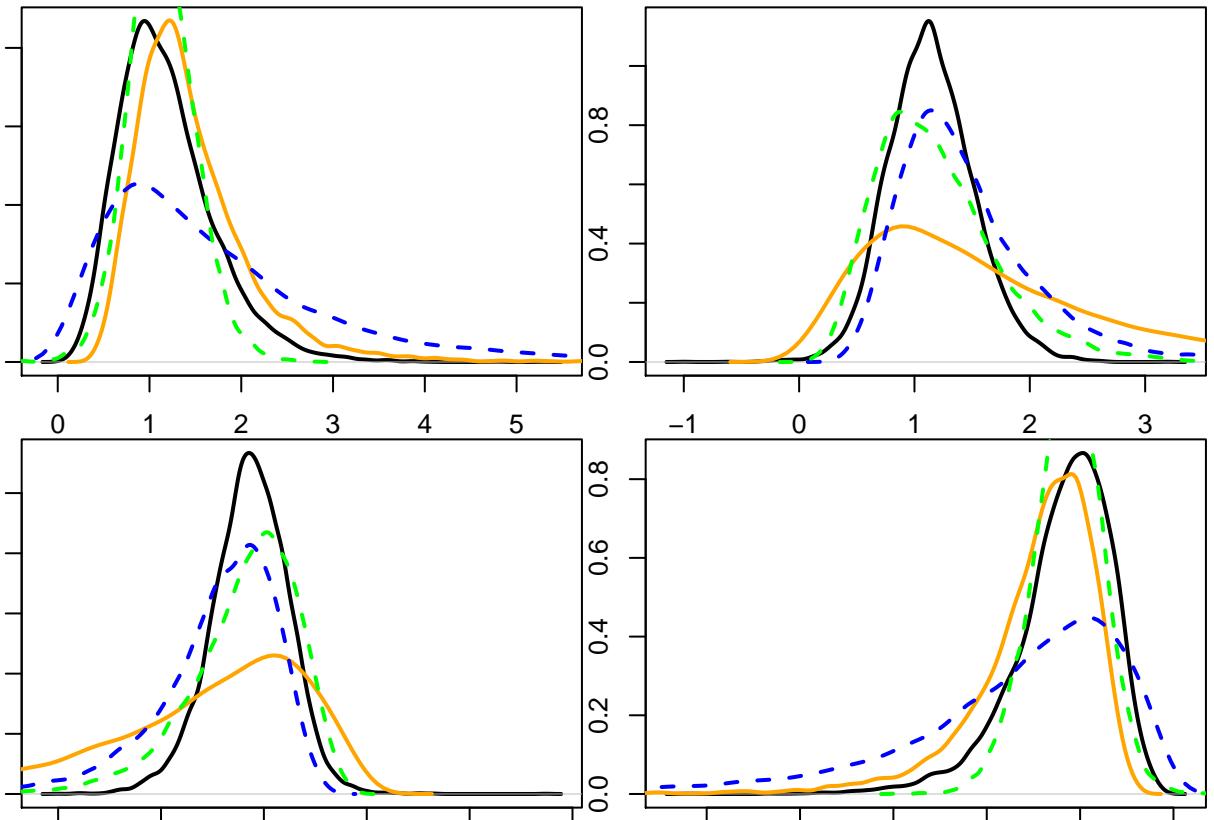


Figure 8. Comparison of Glass's ds when choosing either sd1 or sd2 as standardizer when both samples are extracted from a distribution where sigma=1, m1=1,m2=0, G1=0 and G2=95.75, and either n1=100 and n2=20, or n1=20 and n2=100

¹³² mean difference.



— $\sigma = \text{SD1 (when } n_1 = 100)$
- - - $\sigma = \text{SD2 (when } n_2 = 20)$
- - - $\sigma = \text{SD1 (when } n_1 = 20)$
- - - $\sigma = \text{SD2 (when } n_2 = 100)$



134

135 In the two top plots in Figure ??, $\bar{X}_1 - \bar{X}_2$ is positive. When distributions are right-skewed
 136 (top right), choosing s_2 (i.e. the SD that is negatively correlated with $\bar{X}_1 - \bar{X}_2$) associated
 137 with the smallest sample size (i.e. $n_2 = 20$) will result in the most biased and variable
 138 estimation of glass's d_s . The best estimation will occur when choosing s_1 associated with the
 139 largest sample size (i.e. $n_1 = 100$). On the other side, when distributions are left-skewed (top
 140 left), choosing s_1 (i.e. the SD that is negatively correlated with $\bar{X}_1 - \bar{X}_2$) associated with
 141 the smallest sample size (i.e. $n_1 = 20$) will result in the most biased and variable estimation
 142 of glass's d_s . The less biased and variable glass d_s will occur when s_2 is chosen as
 143 standardizer and $n_2 = 100$.

144 In the two bottom plots in Figure ??, $\bar{X}_1 - \bar{X}_2$ is negative. When distributions are
 145 right-skewed (bottom right), choosing s_1 (i.e. the SD that is positively correlated with
 146 $\bar{X}_1 - \bar{X}_2$) associated with the smallest sample size (i.e. $n_1 = 20$) will result in the most
 147 biased and variable estimation of glass's d_s . The best estimation will occur when choosing s_2

148 associated with the largest sample size (i.e. $n_2 = 100$). On the other side, when distributions
149 are left-skewed (top left), choosing s_2 (i.e. the SD that is positively correlated with $\bar{X}_1 - \bar{X}_2$)
150 associated with the smallest sample size (i.e. $n_2 = 20$) will result in the most biased and
151 variable estimation of glass's d_s . The less biased and variable glass d_s will occur when s_1 is
152 chosen as standardizer and $n_1 = 100$.

153 **When two samples are extracted from distributions with identical shapes, with**

154 $\sigma_1 \neq \sigma_2$ **and** $n_1 = n_2$