International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 13, 2013, pp 1-19

Some Further Aspects of Assessment of Agreement involving Bivariate Normal Responses

Ganesh Dutta
Basanti Devi College
147B, Rash Behari Avenue
Kolkata-700029, West Bengal, India

Bikas Kumar Sinha
Retired Faculty, Indian Statistical Institute
Kolkata-700108, India

[Received May 7, 2013; Revised September 29, 2013; Accepted October 2, 2013]

Abstract

There is an impressive published literature on the statistical issue of assessment of agreement among two or more raters involving both qualitative and quantitative data. In this article we have focused only on the data based on a continuous measurement. In such frameworks, there are several usual approaches for evaluating agreement. Applied to data arising out of a bivariate normal distribution (either naturally or under suitable transformation(s)), Lin's method (1989) was further pursued in Yimprayoon et al. (2006) as a multi-parameter testing problem involving the means, variances and the correlation coefficient of the two measurement distributions. Specifically, it was posed as one of testing the composite null hypothesis H_0: $\mu_x = \mu_y$, $\sigma_x = \sigma_y$ and $\rho \geq \rho_0$ [close to 1]. This formulation corresponds to what is referred to as *perfect agreement* scenario when ρ=1 and this is too much to expect in real life situations. In this study, we have developed large sample likelihood ratio test [LRT] for a more meaningful hypothesis of the form $\mathrm{H}_0: \ |\mu_x - \mu_y| \geq \varepsilon_0, \ \frac{\sigma_x}{\sigma_y} \ or \ \frac{\sigma_y}{\sigma_x} \geq \eta_0, \ \rho \leq \rho_0$ where ε_0 is close to zero and η_0 and ρ_0 are close to unity - all are assumed to be specified. We evaluate the performance of the test when X and Y do not agree at the desired levels of ε_0, η_0 and ρ_0.

Keywords and Phrases: Bivariate Normal, Concordance Correlation Coefficient, Continuous Scale, Likelihood Ratio Test.

AMS Classification: Primary 62H12, Secondary 62H15.

1 Introduction

Measurements of agreement are needed to assess the acceptability of new or generic process, methodology, and formulation in both science and non-science fields of laboratory performance, instrument or assay validation, method comparisons, statistical process control, goodness of fit, and individual bioequivalence. For example, the agreement of laboratory measurements collected in various laboratories, the agreement of a newly developed method with gold standard method, the agreement of manufacturing process measurements with specifications, the agreement of observed values with predicted values, and the agreement in bioavailability of a new or generic formulation with a commonly used formulation. By the way, measuring agreement has been used very often to designate the level of agreement between different data-generating sources referred to as observers or raters. A rater could be a chemist, a psychologist, a radiologist, a clinician, a nurse, a rating system, a diagnosis, a treatment, an instrument, a method, a process, a technique or a formula.

Evaluation of agreement has received considerable attention in the literature more than one and a half centuries ago. Cohen (1960, 1968) discussed this problem in the context of categorical data. Bland and Altman (1986) proposed a simple and meaningful graphical approach for assessing the agreement between two clinical measurements. In a series of articles, Lin (1989, 1992, 1997, 2000) and Lin and Torbeck (1998) examined this problem critically in the framework of method reproducibility and suggested a few measures and studied their properties. In the context of bioequivalence, similar studies have been reported by Anderson and Hauck (1990), Sheiner (1992), Holder and Hsuan (1993), Schall and Luus (1993), Schall (1995), Schall and Williams (1996), and Lin (2000). In the context of goodness of fit, Vonesh, Chinchilli, and Pu (1996) and Vonesh and Chinchilli (1997) have modified Lin's approach for choosing models that have better agreement between the observed and the predicted values. A comprehensive account of the methods for studying intra- and inter-rater agreements is available in the latest book in this area by Lin et al. (2012).

In this article, we have focused only on the data for two competing raters measured on a continuous scale. There are several usual approaches for evaluating agreement for such paired data such as Pearson correlation coefficient, regression analysis, paired t-tests, least squares analysis for slope and intercept, within-subject coefficient of variation, and intra-class correlation coefficient. The concordance correlation coefficient (CCC) was first proposed by Lin (1989) for assessment of agreement in continuous data. It represents a breakthrough in assessing agreement between two distinct methods for continuous data in that it appears to avoid all the shortcomings associated with usual approaches in some situations. In short, Lin (1989) expresses the degree of concordance between two variables X and Y by the mean of their squared difference (MSD), $\mathrm{E(X\text{-}Y)}^2$, and defines the CCC as

$$\rho_c = 1 - \frac{\mathrm{E(Y\text{-}X)}^2}{\mathrm{E_{Indep}(Y\text{-}X)}^2} = \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \tag{1}$$

where $\mathrm{E_{Indep}}(.)$ represents expectation under the assumption of independence of X and Y, μ_x=E(X), μ_y =E(Y), σ_x^2 =Var(X), σ_y^2=Var(Y), and σ_{xy}=Cov(X,Y)=$\rho\sigma_x\sigma_y$. Lin (1989) estimates this CCC with data by substituting the sample moments of an independent bivariate sample into above formula to compute the sample counterpart of CCC (r_c). The CCC translates the MSD into a correlation coefficient that measures the agreement along the identity line. It has the properties of a CCC in that it ranges between -1 and +1, with -1 indicating perfect reversed agreement (Y=−X), 0 indicating no agreement, and +1 indicating perfect agreement (Y=X). Lin et al. (2002) gave a review and comparison of various measures, including the CCC, of developments in this field by comparing the powers of the tests: 1) $\mu_x = \mu_y$, 2) $\sigma_x = \sigma_y$, and 3) $\rho = \rho_0$, where ρ_0 is a given value. Their calculation is illustrated using a real data example. This work was further extended in Hedayat et al. (2009) involving multiple raters. In another direction, Yimprayoon et al. (2006) extended the work of Lin et al. (2002) by combining the problems of testing for $\mu_x = \mu_y$, $\sigma_x = \sigma_y$, and $\rho \geq \rho_0$ into one overall testing problem under bivariate normal set-up and then they presented the result based on simulation study. In this article, we have revisited this testing problem and tried to reformulate an appropriate hypothesis by considering probability of the absolute value of X-Y (=D) less than the fixed boundary, κ under bivariate normal set-up. Moreover, we try to find out the appropriate test statistics for this combined testing problem.

2 Construction of Hypothesis

An intuitively clear measurement of agreement is a measure that captures a large proportion of data within a predetermined boundary from target values. In other words, we want the probability of the absolute value of D=Y−X less than the boundary, κ, to be large. This probability is termed in literature as coverage probability (CP) (cf. Lin et al. (2002)) and it is defined as

$$\mathrm{CP}(\kappa) = \mathrm{P}[|D| < \kappa], \tag{2}$$

where X and Y denote random variables representing paired observations for assessing the agreement. We assume that X and Y have a bivariate normal distribution with means μ_x and μ_y, variances σ_x^2 and σ_y^2, correlation coefficient ρ and the covariance of X and Y is $\sigma_{xy} = \rho\sigma_x\sigma_y$. We denote this by

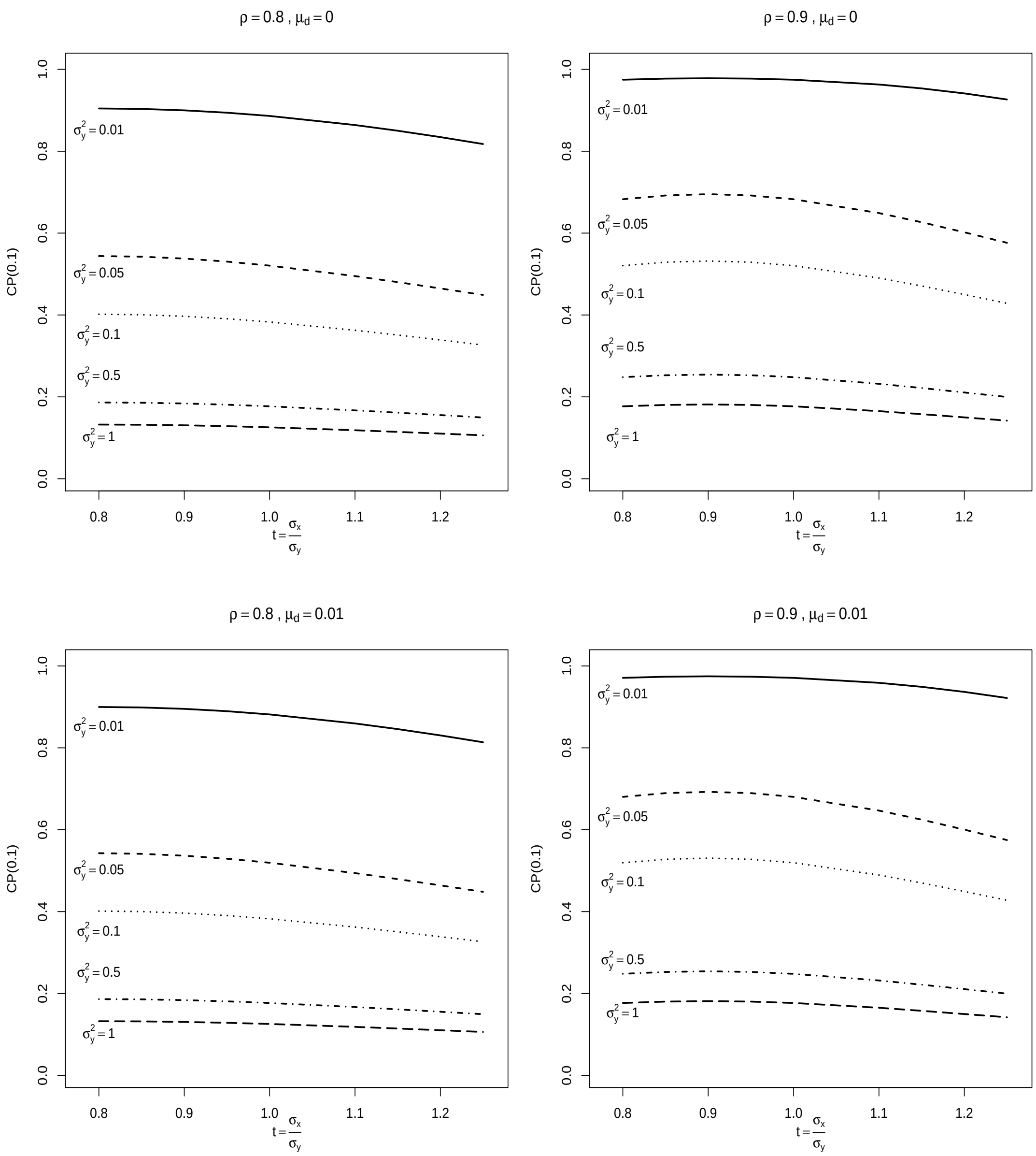
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \ \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\right], \tag{3}$$

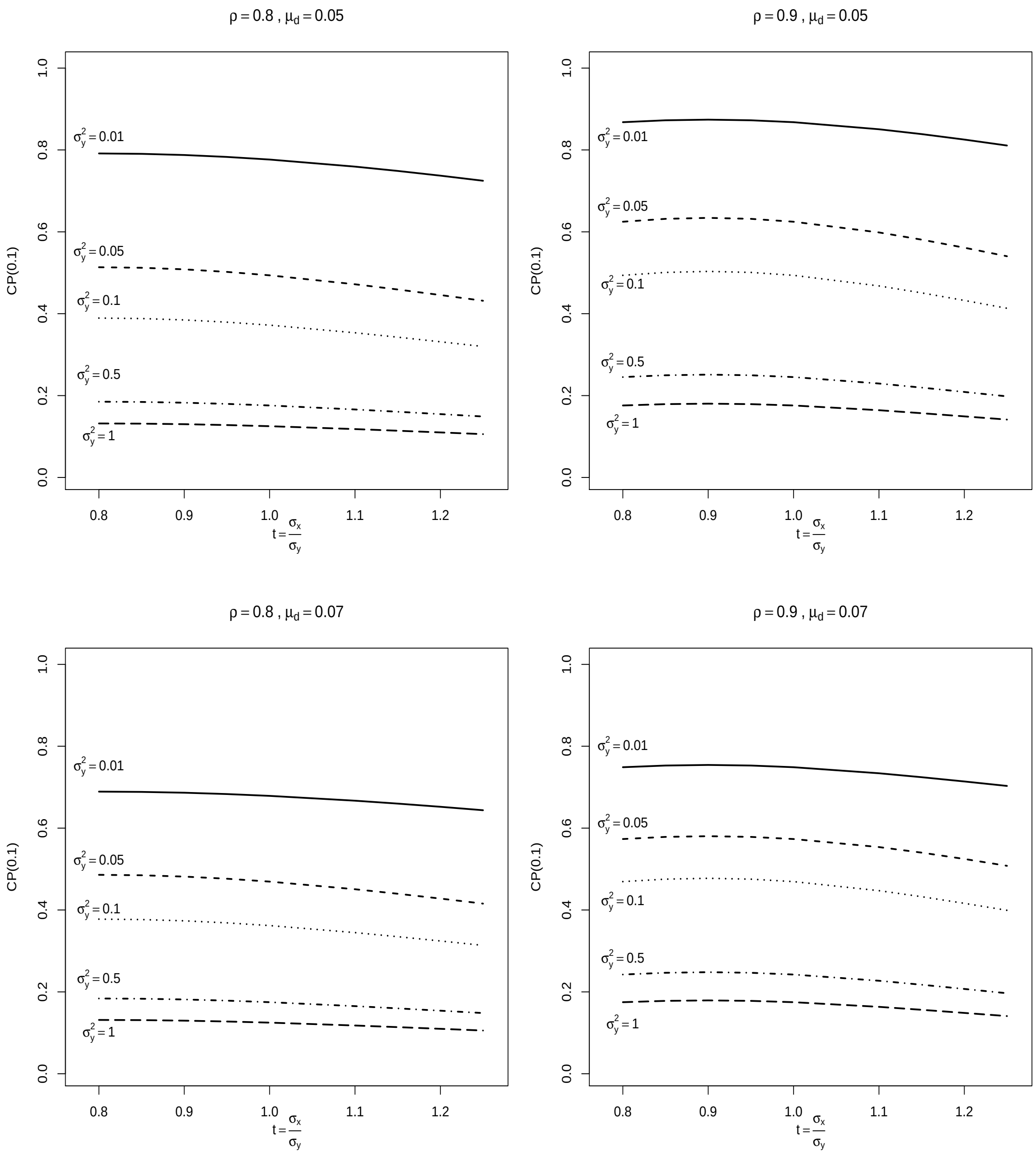
where $-\infty < \mu_x, \ \mu_y < \infty, \ \sigma_x, \ \sigma_y > 0, \ -1 < \rho < 1.$
Under this normality assumption, (2) is reduced to

$$\mathrm{CP}(\kappa) = \mathrm{P}[|D| < \kappa] = \Phi\left(\frac{\kappa - \mu_d}{\sigma_d}\right) - \Phi\left(\frac{-\kappa - \mu_d}{\sigma_d}\right), \tag{4}$$

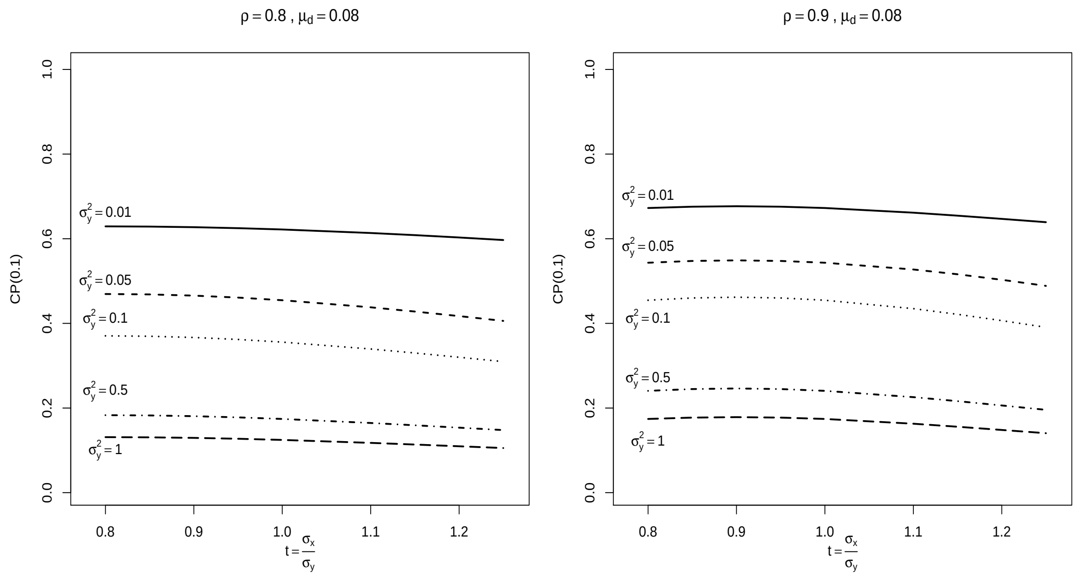
where, $\mu_d = \mu_y - \mu_x$ and $\sigma_d^2 = \sigma_y^2 + \sigma_x^2 - 2\rho\sigma_x\sigma_y = \sigma_y^2(1 + t^2 - 2\rho t)$ and $t = \frac{\sigma_x}{\sigma_y}$.
At this stage, we have critically studied the behaviour of CP for fixed $\kappa = 0.1$ for varying t (=0.8, 0.85, 0.9, 0.95, 1.00, 1.10, 1.15, 1.20, 1.25), μ_d (=0, 0.01, 0.05, 0.07, 0.08), σ_y^2 (=0.01, 0.05, 0.10, 0.50, 1.00) and ρ (=0.8, 0.9).

Figure 1: t vs CP(0, 1)





$\rho = 0.8$, $\mu_d = 0.05$
$\rho = 0.9$, $\mu_d = 0.05$
$\rho = 0.8$, $\mu_d = 0.07$
$\rho = 0.9$, $\mu_d = 0.07$
CP(0.1)
$t = \frac{\sigma_x}{\sigma_y}$
$\sigma_y^2 = 0.01$
$\sigma_y^2 = 0.05$
$\sigma_y^2 = 0.1$
$\sigma_y^2 = 0.5$
$\sigma_y^2 = 1$
0.0
0.2
0.4
0.6
0.8
1.0
0.8
0.9
1.0
1.1
1.2



From Figure 1, it has been observed that CP(0.1) is greater than 0.6 when $0.8 \leq t \leq 1.10$, μ_d=0, 0.01, 0.05, 0.07, 0.08 and $\sigma_y^2 = 0.01$. Therefore from the above observations, it can be concluded that CP for given κ is reasonably satisfactory for $|\mu_x - \mu_y| \leq \varepsilon_0$, $\frac{1}{\eta_0} \leq \frac{\sigma_x}{\sigma_y} \leq \eta_0$, $\rho \geq \rho_0$, where $\varepsilon_0,\ \eta_0 > 0$ are suitably chosen in advance, being close to 0 and 1 respectively.

Therefore, a more appropriate and plausible null hypothesis can be formulated as

$$\mathrm{H}_0:\ |\mu_x - \mu_y| \geq \varepsilon_0,\ \frac{\sigma_x}{\sigma_y}\ or\ \frac{\sigma_y}{\sigma_x} \geq \eta_0,\ \rho \leq \rho_0 \tag{5}$$

where ε_0 is close to zero and η_0 and ρ_0 are close to unity - all are assumed to be specified. We evaluate the performance of the test when X and Y do not agree at the desired levels of ε_0, η_0 and ρ_0. We offer below a solution based on the Likelihood Ratio Test under a bivariate normal set-up. It may be noted that in agreement studies involving two raters, the primary goal of an experimenter is to offer a test procedure which points towards the agreement in a logical sense. That is why, the formulation in Lin et al. (2002) was generalized in Yimprayoon et al. (2006) and that is further generalized in the above. Following Lin et al. (2002), we notice that when there is a disagreement between the two marginal distributions, the source is defined as constant and/or scale "shift", or lack of "accuracy". When there is a disagreement due to large within-sample variation, the source is defined as lack of "precision". The alternative to H_0 formulated above addresses both these issues.

Since the density function for bivariate normal distribution belongs to monotone likelihood ratio family, the testing of composite null hypothesis (5) will be the same as testing the *union* of following four composite hypotheses:

$$\mathrm{H}_{01}: \ \mu_x = \mu, \ \mu_y = \mu + \varepsilon_0, \ \sigma_x = \sigma, \ \sigma_y = \sigma\eta_0, \ \rho = \rho_0 \tag{6}$$

$$\mathrm{H}_{02}: \ \mu_x = \mu, \ \mu_y = \mu + \varepsilon_0, \ \sigma_x = \sigma, \ \sigma_y = \frac{\sigma}{\eta_0}, \ \rho = \rho_0 \tag{7}$$

$$\mathrm{H}_{03}: \ \mu_x = \mu, \ \mu_y = \mu - \varepsilon_0, \ \sigma_x = \sigma, \ \sigma_y = \sigma\eta_0, \ \rho = \rho_0 \tag{8}$$

$$\mathrm{H}_{04}: \ \mu_x = \mu, \ \mu_y = \mu - \varepsilon_0, \ \sigma_x = \sigma, \ \sigma_y = \frac{\sigma}{\eta_0}, \ \rho = \rho_0. \tag{9}$$

It may be noted that the four composite hypotheses stated above point to the four directions of disagreement. Clearly, in working out the over-all likelihood under the union of these four component hypotheses, we will be guided by the five statistics based on n paired observations and these are $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \ S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, \ S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \ S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$.

3 Derivation of Test Procedures

The likelihood function can be written as

$$\begin{aligned}\mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data}) = \frac{1}{(2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^n} \exp\Bigg[&- \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \Bigg\{ \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2 \\ &- 2\rho\left(\frac{x_i - \mu_x}{\sigma_x}\right)\left(\frac{y_i - \mu_y}{\sigma_y}\right) + \left(\frac{y_i - \mu_y}{\sigma_y}\right)^2 \Bigg\}\Bigg].\end{aligned} \tag{10}$$

Our approach will be to work out maximum of the likelihood under each of the four hypotheses stated above and then to compute the largest of the four expressions, so derived, to finally arrive at the numerator of the LRT. We assume without any loss of generality that $\varepsilon_0 > 0, \ \eta_0 > 1$. Next note that the domain of variation of the four statistics viz., $\bar{x}, \bar{y}, S_x \ (= \sqrt{S_{xx}}), S_y \ (= \sqrt{S_{yy}})$ can be logically partitioned as : $D1 : [\bar{x} > \bar{y}; S_x > S_y]; D2 : [\bar{x} > \bar{y}; S_x < S_y]; D3 : [\bar{x} < \bar{y}; S_x > S_y]; D4 : [\bar{x} < \bar{y}; S_x < S_y]$. These will presumably provide largest of the four likelihoods in a logical and expected manner. Under the assumption of $\varepsilon_0 > 0, \ \eta_0 > 1$, it should turn out that $D4$ favors H_{01}, $D3$ favors H_{02}, $D2$ favors H_{03} and $D1$ favors H_{04}. This is readily verified to be true. Details are shown in the Appendix.

We use the statistic λ_1^* from the LRT defined by

$$\lambda_1 = \frac{\max\limits_{\Theta_{01}} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data})}{\max\limits_{\Theta} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data})}, \tag{11}$$

where $\Theta = (\mu_x,\ \mu_y,\ \sigma_x,\ \sigma_y,\ \rho)$ and $\Theta_{01} = (\mu,\ \sigma)$.

We reject the composite hypotheses if

$$\lambda_1 < d_1^*. \tag{12}$$

To evaluate $\max\limits_{\Theta_{01}} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data})$ it can be easily shown that when we confine to the domain $D4$, maximum likelihood estimators of μ and σ are as follows:

$$\widehat{\mu} = a\bar{x} + (1-a)\bar{y^*}, \tag{13}$$

$$\widehat{\sigma^2} = \frac{\mathrm{Q}(\widehat{\mu})}{2n(1-\rho_0^2)}, \tag{14}$$

where

$$\bar{y^*} = \bar{y} - \varepsilon_0, \tag{15}$$

$$a = \frac{\eta_0(\eta_0 - \rho_0)}{k}, \tag{16}$$

$$k = 1 + \eta_0^2 - 2\rho_0\eta_0 \tag{17}$$

$$\mathrm{Q}(\mu) = \mathrm{Q}_1(\mu) + \mathrm{Q}_2 \tag{18}$$

$$\mathrm{Q}_1(\mu) = n[(\bar{x} - \mu)^2 - 2\frac{\rho_0}{\eta_0}(\bar{x} - \mu)(\bar{y^*} - \mu) + \frac{1}{\eta_0^2}(\bar{y^*} - \mu)^2)], \tag{19}$$

$$\mathrm{Q}_2 = \mathrm{S}_{xx} - 2\frac{\rho_0}{\eta_0}\mathrm{S}_{xy} + \frac{1}{\eta_0^2}\mathrm{S}_{yy}. \tag{20}$$

Substituting the above estimators, we get

$$\max_{\Theta_{01}} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data}) = \left(\frac{n\sqrt{1-\rho_0^2}}{\pi\eta_0}\right)^n \exp(-n)\mathrm{Q}^{-n}(\widehat{\mu}). \tag{21}$$

Likewise, for all other domains, maximum of the likelihood has been displayed in the Appendix.

To find $\max\limits_{\Theta} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data})$, it is well known that $\widehat{\widehat{\mu_x}} = \bar{x}$, $\widehat{\widehat{\mu_y}} = \bar{y}$, $\widehat{\widehat{\sigma_x^2}} = \frac{\mathrm{S}_{xx}}{n}$, $\widehat{\widehat{\sigma_y^2}} = \frac{\mathrm{S}_{yy}}{n}$, $\widehat{\widehat{\rho}} = \frac{\mathrm{S}_{xy}}{\mathrm{S}_{xx}\mathrm{S}_{yy}} = r$, where r is sample correlation coefficient, which gives

$$\max_{\Theta} \mathrm{L}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho|\text{data}) = \frac{n^n \exp(-n)}{\left(\sqrt{\mathrm{S}_{xx}\mathrm{S}_{yy} - \mathrm{S}_{xy}^2}\right)^n (2\pi)^n}. \tag{22}$$

Hence in case of domain $D4$, λ_1 is obtained as

$$\lambda_1 = \left(\frac{2\sqrt{1-\rho_0^2}}{\eta_0}\right)^n \left(\frac{\sqrt{\mathrm{Q}_2^*}}{\mathrm{Q}_1(\widehat{\mu}) + \mathrm{Q}_2}\right)^n, \tag{23}$$

where

$$Q_2^* = S_{xx}S_{yy} - S_{xy}^2. \tag{24}$$

Now we have seen that under H_{01},

$$\frac{\sqrt{n}(\bar{x} - \bar{y}^*)}{\sigma\sqrt{k}} \sim N(0, 1), \tag{25}$$

where $E_{H_{01}}(\bar{x} - \bar{y}^*) = 0$ and $\mathrm{Var}_{H_{01}}(\bar{x} - \bar{y}^*) = \frac{\sigma^2 k}{n}$.
Again

$$Q_1(\widehat{\mu}) = nk_1(\bar{x} - \bar{y}^*)^2, \tag{26}$$

where

$$k_1 = (1 - 2a) + \frac{2a\rho_0\eta_0 + a^2 k}{\eta_0^2}. \tag{27}$$

It is also known that $\begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix} \sim W(\Sigma,\ n-1)$ where W represents Wishart distribution and $\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$. It is also known that E(S_{ij})=(n-1)σ_{ij} and Cov($S_{ij},\ S_{kl}$)=(n-1)$(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$. Hence under H_{01}, E(Q_2)=2(n-1)$\sigma^2(1-\rho_0^2)$, E(Q_2^*)=(n-1)(n-2)$\sigma^4\eta_0^2(1-\rho_0^2)$ and

$$\frac{Q_2}{\sqrt{Q_2^*}} \xrightarrow{P} \frac{2\sqrt{1-\rho_0^2}}{\eta_0}. \tag{28}$$

Likewise, combining all the domains, we find that the test depends on the two sample means only through the absolute value of their difference. Hence, under the union of the component hypotheses,

$$\begin{aligned}
\lim_{n\to\infty} P[\lambda_1 < d_1^*] = \lim_{n\to\infty} P[T_1 > d_1] &= \lim_{n\to\infty} P\left[\frac{Q_1(\widehat{\mu})}{(Q_2^*)^{\frac{1}{2}}} > d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}\right] \\
&= \lim_{n\to\infty} P\left[\frac{\sqrt{nk_1}|\bar{x} - \bar{y}^*|}{(Q_2^*)^{\frac{1}{4}}} > \left(d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}\right)^{\frac{1}{2}}\right] \\
&= \lim_{n\to\infty} P\left[\frac{\sqrt{n}|\bar{x} - \bar{y}^*|}{(k^2 Q_2^*)^{\frac{1}{4}}} > \left(\frac{d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}}{k_1 k}\right)^{\frac{1}{2}}\right] \\
&= \lim_{n\to\infty} P\left[\frac{|V|}{(V_1^2 V_2^2)^{\frac{1}{4}}} > \eta_0^{\frac{1}{2}}(1-\rho_0^2)^{\frac{1}{4}}\left(\frac{d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}}{k_1 k}\right)^{\frac{1}{2}}\right] \\
&= \lim_{n\to\infty} P\left[|T| > \eta_0^{\frac{1}{2}}(1-\rho_0^2)^{\frac{1}{4}}\left(\frac{d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}}{k_1 k}\right)^{\frac{1}{2}}\right]
\end{aligned}$$

(29)

where $V=\frac{\sqrt{n}(\bar{x}-\bar{y^*})}{\sigma\sqrt{k}}$, $V_1^2V_2^2=\frac{Q_2^*}{\sigma^4\eta_0^2(1-\rho_0^2)}$, $T=\frac{V}{(V_1^2V_2^2)^{\frac{1}{4}}}$
and

$$V\sim N(0,1),\quad V_1^2\sim\ \chi^2_{n-1},\qquad V_2^2\sim\ \chi^2_{n-2}\text{under }\ H_{01} \tag{30}$$

and V, V_1 and V_2 are independently distributed.

3.1 Limiting distribution of T

The joint distribution of (V, V_1^2, V_2^2) is

$$f(v,v_1^2,v_2^2)=\frac{1}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})\Gamma(\frac{1}{2})2^{n-1}}\exp\Bigg[-\frac{v^2+v_1^2+v_2^2}{2}\Bigg]v_1^{n-3}v_2^{n-4};\quad -\infty<v<\infty,\ v_1,v_2>0. \tag{31}$$

The joint distribution of (V, V_1, V_2) is

$$f(v,v_1,v_2)=\frac{1}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})\Gamma(\frac{1}{2})2^{n-3}}\exp\Bigg[-\frac{v^2+v_1^2+v_2^2}{2}\Bigg]v_1^{n-2}v_2^{n-3};\quad -\infty<v<\infty,\ v_1,v_2>0. \tag{32}$$

Transform (V, V_1, V_2)$\longrightarrow$ (R,θ_1,θ_2), where

$$\begin{aligned} v&=R\cos\theta_1\\ v_1&=R\sin\theta_1\cos\theta_2\\ v_2&=R\sin\theta_1\sin\theta_2;\qquad 0<\theta_1<\pi,\ 0<\theta_2<\frac{\pi}{2}. \end{aligned}$$

The Jacobian of the transformation is

$$\frac{\partial(v,v_1,v_2)}{\partial(R,\theta_1,\theta_2)}=\begin{vmatrix}\cos\theta_1 & \sin\theta_1\cos\theta_2 & \sin\theta_1\sin\theta_2\\ -R\sin\theta_1 & R\cos\theta_1\cos\theta_2 & R\cos\theta_1\sin\theta_2\\ 0 & -R\sin\theta_1\sin\theta_2 & R\sin\theta_1\cos\theta_2\end{vmatrix}=R^2\sin\theta_1; \tag{33}$$

and

$$v^2+v_1^2+v_2^2=R^2. \tag{34}$$

Therefore the joint distribution of (R, θ_1, θ_2) is

$$f(R,\theta_1,\theta_2)=\frac{1}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})\Gamma(\frac{1}{2})2^{n-3}}\exp\Bigg[-\frac{R^2}{2}\Bigg]R^{2n-3}(\sin\theta_1)^{2n-4}(\sin\theta_2)^{n-3}(\cos\theta_2)^{n-2};$$

$$R>0,\ 0<\theta_1<\pi,\ 0<\theta_2<\frac{\pi}{2}. \tag{35}$$

We have seen from Expression (35) that R,θ_1,θ_2 are independent.

The joint distribution of $(\theta_1,\ \theta_2)$

$$f(\theta_1,\theta_2) = \frac{2}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}(\sin\theta_1)^{2n-4}(\sin\theta_2)^{n-3}(\cos\theta_2)^{n-2};$$

$$0<\theta_1<\pi,\ 0<\theta_2<\frac{\pi}{2}. \qquad (36)$$

Transform $(\theta_1,\ \theta_2)\longrightarrow(\mathrm{T},\ \theta_2)$,

where $t = \frac{\cot\theta_1}{\sqrt{\sin\theta_2\cos\theta_2}}; \quad -\infty<t<\infty.$

Jacobian of transformation is,

$$\frac{\partial\theta_1}{\partial t} = \frac{\sqrt{\sin\theta_2\cos\theta_2}}{1+t^2\sin\theta_2\cos\theta_2}.$$

Therefore the joint distribution of $(\mathrm{T},\ \theta_2)$ is

$$f(t,\theta_2) = \frac{2}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})} \times \frac{(\sin\theta_2)^{n-\frac{5}{2}}(\cos\theta_2)^{n-\frac{3}{2}}}{(1+t^2\sin\theta_2\cos\theta_2)^{n-1}}$$

$$= \frac{2}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}\sum_{j=0}^{\infty}\binom{-(n-1)}{j}t^{2j}(\sin\theta_2)^{n+j-\frac{5}{2}}(\cos\theta_2)^{n+j-\frac{3}{2}}. \qquad (37)$$

Then the marginal distribution of T is

$$f(t) = \frac{2}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}\sum_{j=0}^{\infty}\binom{-(n-1)}{j}t^{2j}\int_0^{\frac{\pi}{2}}(\sin\theta_2)^{n+j-\frac{5}{2}}(\cos\theta_2)^{n+j-\frac{3}{2}}d\theta_2$$

$$= \frac{1}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}\sum_{j=0}^{\infty}\binom{-(n-1)}{j}\mathrm{B}\left(\frac{n+j-\frac{3}{2}}{2},\frac{n+j-\frac{1}{2}}{2}\right)t^{2j}$$

$$= \frac{1}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}t^{2j}\frac{(n+j-2)!}{(n-2)!}\mathrm{B}\left(\frac{n+j-\frac{3}{2}}{2},\frac{n+j-\frac{1}{2}}{2}\right). \qquad (38)$$

Now we recall Stirling's approximation

$$\mathrm{B}(x,y) \simeq \sqrt{2\pi}\frac{x^{x-\frac{1}{2}}y^{y-\frac{1}{2}}}{(x+y)^{x+y-\frac{1}{2}}} \quad \text{for large } x \text{ and } y;$$

$$\mathrm{B}(x,y) \simeq \Gamma(y)x^{-y} \quad \text{for large } x \text{ but } y \text{ is fixed;}$$

$$n! \simeq \sqrt{2\pi}n^{n+\frac{1}{2}}\exp(-n).$$

$$\frac{(n+j-2)!}{(n-2)!}\times\frac{\mathrm{B}\left(\frac{n+j-\frac{3}{2}}{2},\frac{n+j-\frac{1}{2}}{2}\right)}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})}\simeq\frac{\sqrt{2\pi}(n+j-2)^{n+j-2+\frac{1}{2}}\exp(-(n+j-2))}{\sqrt{2\pi}(n-2)^{n-2+\frac{1}{2}}\exp(-(n-2))}\times$$

$$\frac{\sqrt{2\pi}\frac{\left(\frac{n+j-\frac{3}{2}}{2}\right)^{\frac{n+j-\frac{3}{2}}{2}-\frac{1}{2}}\left(\frac{n+j-\frac{1}{2}}{2}\right)^{\frac{n+j-\frac{1}{2}}{2}-\frac{1}{2}}}{\left(\frac{n+j-\frac{3}{2}}{2}+\frac{n+j-\frac{1}{2}}{2}\right)^{\frac{n+j-\frac{3}{2}}{2}+\frac{n+j-\frac{1}{2}}{2}-\frac{1}{2}}}}{\Gamma(\frac{1}{2})(n-\frac{3}{2})^{-\frac{1}{2}}\times\sqrt{2\pi}\frac{\left(\frac{n-2}{2}\right)^{\frac{n-2}{2}-\frac{1}{2}}\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}-\frac{1}{2}}}{\left(\frac{n-2}{2}+\frac{n-1}{2}\right)^{\frac{n-2}{2}+\frac{n-1}{2}-\frac{1}{2}}}}$$

$$=\left(1+\frac{j}{n-2}\right)^{n-2+\frac{1}{2}}n^j\left(1+\frac{j}{n-2}\right)^{j}\exp(-j)\times$$

$$\frac{\frac{(\frac{n}{2})^{\frac{n+j-\frac{3}{2}}{2}-\frac{1}{2}+\frac{n+j-\frac{1}{2}}{2}-\frac{1}{2}}\left(1+\frac{j-\frac{3}{2}}{n}\right)^{\frac{n+j-\frac{3}{2}}{2}-\frac{1}{2}}\left(1+\frac{j-\frac{1}{2}}{n}\right)^{\frac{n+j-\frac{1}{2}}{2}-\frac{1}{2}}}{n^{n+j-\frac{3}{2}}\left(1+\frac{j-1}{n}\right)^{n+j-\frac{3}{2}}}}{\sqrt{\pi}n^{-\frac{1}{2}}(1-\frac{3}{2n})^{-\frac{1}{2}}\times\frac{(\frac{n}{2})^{\frac{n-2}{2}+\frac{n-1}{2}-1}(1-\frac{2}{n})^{\frac{n-2}{2}-\frac{1}{2}}(1-\frac{1}{n})^{\frac{n-1}{2}-\frac{1}{2}}}{n^{\frac{n-2}{2}+\frac{n-1}{2}-\frac{1}{2}}(1-\frac{3}{2n})^{n-2}}}$$

$$\simeq\frac{\exp(j)n^{j+\frac{1}{2}}\exp(-j)\times\frac{\frac{1}{2^{n+j-2}}\times\exp(\frac{j-\frac{3}{2}}{2}+\frac{j-\frac{1}{2}}{2})\times n^{n+j-2}}{n^{n+j-\frac{3}{2}}\times\exp(j-1)}}{\frac{\sqrt{\pi}\frac{n^{n-\frac{5}{2}}}{2^{n-\frac{5}{2}}}\exp\left(-\frac{3}{2}\right)}{n^{n-2}\exp(-\frac{3}{2})}}$$

$$=\frac{\sqrt{n}}{\sqrt{2\pi}}\frac{n^j}{2^j}.$$

Here we use this fact:

$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n=\exp(a).$$

Therefore the limiting distribution of T is

$$f(t)\simeq\frac{\sqrt{n}}{\sqrt{2\pi}}\sum_{j=0}^{\infty}(-1)^j\frac{(\frac{nt^2}{2})^j}{j!}$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi}} \exp(-\frac{nt^2}{2}).$$

Hence under H_{01}, $\sqrt{n}\mathrm{T} \xrightarrow{a} \mathrm{N}(0,\ 1)$.

4 Computation of the cut-off points in small samples

In this section we briefly illustrate computation of d_1 when the sample size n is small. Let, under H_{01},

$$\mathrm{P}[|T| > t] = \alpha \tag{39}$$

that is

$$\int_{-t}^{t}\int_{0}^{\frac{\pi}{2}} f(T,\theta_2)d\theta_2 dT = 1-\alpha \tag{40}$$

where,

$$f(T,\theta_2) = \frac{2}{\mathrm{B}(n-\frac{3}{2},\frac{1}{2})\mathrm{B}(\frac{n-2}{2},\frac{n-1}{2})} \times \frac{(\sin\theta_2)^{n-\frac{5}{2}}(\cos\theta_2)^{n-\frac{3}{2}}}{(1+T^2\sin\theta_2\cos\theta_2)^{n-1}} \tag{41}$$

$$t = \eta_0^{\frac{1}{2}}(1-\rho_0^2)^{\frac{1}{4}}\left(\frac{d_1 - \frac{2\sqrt{1-\rho_0^2}}{\eta_0}}{k_1 k}\right)^{\frac{1}{2}}. \tag{42}$$

Let

$$\mathrm{I} = \int_{-t}^{t}\int_{0}^{\frac{\pi}{2}} f(T,\theta_2)d\theta_2 dT. \tag{43}$$

Here we use Monte Carlo integration to find I. For this, we generate separately s independent uniform random variables T_1, T_2,...,T_s on the interval $[-t,\ t]$ and s independent uniform random variables $\theta_{21},\ \theta_{22},...,\theta_{2s}$ on the interval $[0,\ \frac{\pi}{2}]$ and compute

$$\widehat{\mathrm{I}} = \frac{\pi t}{s^2}\sum_{i=1}^{s}\sum_{j=1}^{s} f(T_i,\theta_{2i}). \tag{44}$$

Step 1: First we find initial value t_0 of t by trial and error method such that I$\simeq 1-\alpha$, for given α.
Step 2: Compute $\widehat{\mathrm{I}}$ for t_i where $t_i = t_0 + di$, $i = 1, 2, ..., l$
Step 3: Repeat Step 2 based on 100 simulations.
Step 4: Repeat Steps 1 through 4 using $\rho_0 = 0.7,\ 0.8,\ 0.9,\ \eta_0 = 0.8,\ 1,\ 1.25$ and $n = 5,\ 10,\ 15,\ 20,\ 100$.
Step 5: Find values of d_1 for the level α=1%, 5%.
The simulated cut-off points are shown in Table 1 for $s = 1000,\ l = 100,\ d = 0.0001$.

Table 1: Cut-off points

$\alpha = 0.05$

η_0	ρ_0	n	T	d_1
0.8	0.7	5	1.3717	3.464986
		10	0.7554	2.294746
		15	0.5622	2.067505
		20	0.4755	1.987192
		100	0.1950	1.819301
	0.8	5	1.3717	2.911171
		10	0.7554	1.927972
		15	0.5622	1.737052
		20	0.4755	1.669575
		100	0.1950	1.528519
	0.9	5	1.3717	2.114916
		10	0.7554	1.400639
		15	0.5622	1.261939
		20	0.4755	1.212918
		100	0.1950	1.110443
1	0.7	5	1.3717	2.771989
		10	0.7554	1.835796
		15	0.5622	1.654004
		20	0.4755	1.5897536
		100	0.1950	1.4554410
	0.8	5	1.3717	2.328937
		10	0.7554	1.542377
		15	0.5622	1.389641
		20	0.4755	1.3356601
		100	0.1950	1.2228150
	0.9	5	1.3717	1.691933
		10	0.7554	1.120511
		15	0.5622	1.009551
		20	0.4755	0.9703346
		100	0.1950	0.8883545
1.25	0.7	5	1.3717	2.217591
		10	0.7554	1.468637
		15	0.5622	1.3232032
		20	0.4755	1.2718028
		100	0.1950	1.1643528
	0.8	5	1.3717	1.863149
		10	0.7554	1.233902
		15	0.5622	1.1117130
		20	0.4755	1.0685281
		100	0.1950	0.9782520
	0.9	5	1.3717	1.353547
		10	0.7554	0.896409
		15	0.5622	0.8076408
		20	0.4755	0.7762677
		100	0.1950	0.7106836

$\alpha = 0.01$

η_0	ρ_0	n	T	d_1
0.8	0.7	5	2.3554	6.737857
		10	1.3633	3.444478
		15	0.7307	2.261978
		20	0.6871	2.206796
		100	0.4873	1.997334
	0.8	5	2.3554	5.660932
		10	1.3633	2.893940
		15	0.7307	1.900442
		20	0.6871	1.854080
		100	0.4873	1.678096
	0.9	5	2.3554	4.112572
		10	1.3633	2.102399
		15	0.7307	1.380639
		20	0.6871	1.346958
		100	0.4873	1.219108
1	0.7	5	2.3554	5.390285
		10	1.3633	2.755582
		15	0.7307	1.809583
		20	0.6871	1.765437
		100	0.4873	1.5978670
	0.8	5	2.3554	4.528745
		10	1.3633	2.315152
		15	0.7307	1.520353
		20	0.6871	1.483264
		100	0.4873	1.3424768
	0.9	5	2.3554	3.290057
		10	1.3633	1.681919
		15	0.7307	1.104511
		20	0.6871	1.077566
		100	0.4873	0.9752868
1.25	0.7	5	2.3554	4.312228
		10	1.3633	2.204466
		15	0.7307	1.447666
		20	0.6871	1.412350
		100	0.4873	1.2782936
	0.8	5	2.3554	3.622996
		10	1.3633	1.852122
		15	0.7307	1.216283
		20	0.6871	1.186611
		100	0.4873	1.0739814
	0.9	5	2.3554	2.632046
		10	1.3633	1.345535
		15	0.7307	0.883609
		20	0.6871	0.862053
		100	0.4873	0.7802294

5 Derivation of cut-off points for large sample size

In this section we discuss the derivation of cut-off points for large sample size. The cut-off points under normal approximation are shown in Table 2 for $n = 15,\ 20,\ 100,\ \eta_0 = 0.8,\ 1,\ 1.25,\ \rho_0 = 0.7,\ 0.8,\ 0.9$.

Table 2: Cut-off points under normality approximation

$\alpha = 0.05$

n	$\tau_{\alpha/2}$	η_0	ρ_0	d_1
15	2.575829	0.8	0.7	1.811681
			0.8	1.522116
			0.9	1.105792
		1	0.7	1.4493446
			0.8	1.2176931
			0.9	0.8846335
		1.25	0.7	1.1594757
			0.8	0.9741544
			0.9	0.7077068
20	2.575829	0.8	0.7	1.800164
			0.8	1.512440
			0.9	1.098762
		1	0.7	1.440131
			0.8	1.209952
			0.9	0.879010
		1.25	0.7	1.1521051
			0.8	0.9679619
			0.9	0.7032080
100	2.575829	0.8	0.7	1.844585
			0.8	1.549762
			0.9	1.125876
		1	0.7	1.4756683
			0.8	1.2398094
			0.9	0.9007006
		1.25	0.7	1.1805347
			0.8	0.9918475
			0.9	0.7205605

$\alpha = 0.01$

n	$\tau_{\alpha/2}$	η_0	ρ_0	d_1
15	1.959964	0.8	0.7	1.800598
			0.8	1.512805
			0.9	1.099027
		1	0.7	1.4404784
			0.8	1.2102439
			0.9	0.8792218
		1.25	0.7	1.1523827
			0.8	0.9681951
			0.9	0.7033774
20	1.959964	0.8	0.7	1.793930
			0.8	1.507203
			0.9	1.094957
		1	0.7	1.4351441
			0.8	1.2057622
			0.9	0.8759659
		1.25	0.7	1.1481152
			0.8	0.9646098
			0.9	0.7007727
100	1.959964	0.8	0.7	1.819649
			0.8	1.528811
			0.9	1.110655
		1	0.7	1.4557192
			0.8	1.2230488
			0.9	0.8885243
		1.25	0.7	1.1645754
			0.8	0.9784390
			0.9	0.7108195

Acknowledgement:
This study was initiated during one of the visits of the second author to the North-East Centre of the Indian Statistical Institute in Tezpur, India. We are highly thankful to

Professor S. M. Bendre, Head of the Centre, for providing excellent academic/research atmosphere during the visits of the second author. We also thank Dr. Shalini Chandra for her participation in the early discussions on this topic. First author and Dr. Chandra were attached to the centre during that time.

The authors acknowledge constructive comments from the referee.

References

[1] Anderson, S. and Hauck, W. W. (1990). Consideration of individual bioequivalence, *Journal of pharmacokinetics and biopharmaceutics*, **18**, 259-273.

[2] Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*, John wiley and sons, New York.

[3] Bland, J. M. and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement, *The Lancet*, **8**, 307-310.

[4] Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and psychological measurement*, **20**, 37-46.

[5] Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit, *Psychological bulletin*, **70**, 213-220.

[6] Hedayat, A. S., Lou, C. and Sinha, Bikas K. (2009). A statistical approach to assessment of agreement involving multiple raters, *Communications in statistics - theory & methods*, **38**, 2899 - 2922.

[7] Holder, D. J. and Hsuan, F. (1993). Moment-based criteria for determining bioequivalence, *Biometrika*, **80**, 835-846.

[8] Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility, *Biometrics*, **45**, 255-268.

[9] Lin, L. I. (1992). Assay validation using the concordance correlation coefficient, *Biometrics*, **48**, 599-604.

[10] Lin, L. I. (1997). Rejoinder to the letter to the editor by Atkinson and Nevill, *Biometrics*, **53**, 777-778.

[11] Lin, L. I. (2000). Total deviation index for measuring individual agreement: with application in lab performance and bioequivalence, *Statistics in medicine*, **19**, 255-270.

[12] Lin, L., Hedayat, A. S., and Wu, W. (2012). *Statistical Tools for Measuring Agreement*. Springer.

[13] Lin, L. I., and Torbeck, L. D. (1998). Coefficient of accuracy and concordance correlation coefficient: new statistics for method comparison, *PDA journal of pharmaceutical science and technology*, **52**, 55-59.

[14] Lin, L., Hedayat, A. S., Sinha, B. K. and Yang, M. (2002). Statistical methods in assessing agreement: models, issues, and tools, *Journal of the American statistical association*, **97**, 257-270.

[15] Schall, R. (1995). Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar, *Biometrics*, **51**, 615-626.

[16] Schall, R. and Luus, H. G. (1993). On population and individual bioequivalence, *Statistics in medicine*, **12**, 1109-1124.

[17] Schall, R. and Williams, R. L. (1996). Towards a practical strategy for assessing individual bioequivalence, *Journal of pharmacokinetics and biopharmaceutics*, **24**, 133-149.

[18] Sheiner, L. B. (1992). Bioequivalence revisited, *Statistics in medicine*, **11**, 1777-1788.

[19] Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*, Marcel Dekker, New York.

[20] Vonesh, E. F., Chinchilli, V. M. and Pu, K. (1996). Goodness-of-fit in generalised nonlinear mixed-effect models, *Biometrics*, **52**, 572-587.

[21] Yimprayoon P, Tiensuwan M. and Sinha B. K. (2006). Some statistical aspects of assessing agreement: theory and applications, (English summary), Festschrift for Tarmo Pukkila on his 60th birthday, Dep. Math. Stat. Philos. Univ. Tampere, Tampere, 327-346.

APPENDIX

Four hypotheses:

1. $\mathrm{H}_{01} : \mu_x = \mu,\ \mu_y = \mu + \epsilon_0,\ \sigma_x = \sigma,\ \sigma_y = \sigma\eta_0,\ \rho = \rho_0$
2. $\mathrm{H}_{02} : \mu_x = \mu,\ \mu_y = \mu + \epsilon_0,\ \sigma_x = \sigma,\ \sigma_y = \frac{\sigma}{\eta_0},\ \rho = \rho_0$
3. $\mathrm{H}_{03} : \mu_x = \mu,\ \mu_y = \mu - \epsilon_0,\ \sigma_x = \sigma,\ \sigma_y = \sigma\eta_0,\ \rho = \rho_0$
4. $\mathrm{H}_{04} : \mu_x = \mu,\ \mu_y = \mu - \epsilon_0,\ \sigma_x = \sigma,\ \sigma_y = \frac{\sigma}{\eta_0},\ \rho = \rho_0$

and respective maximum likelihoods

1. $\max_{\mathrm{H}_{01}} \mathrm{L} = \left(\frac{n\sqrt{(1-\rho_0^2)}}{\pi\eta_0}\right)^n \exp(-n)\mathrm{Q}^{-n}(\widehat{\mu}_{\mathrm{H}_{01}}) = \mathrm{L}_1,$ say

2. $\max_{\mathrm{H}_{02}} \mathrm{L} = \left(\frac{n\eta_0\sqrt{(1-\rho_0^2)}}{\pi}\right)^n \exp(-n)\mathrm{Q}^{-n}(\widehat{\mu}_{\mathrm{H}_{02}}) = \mathrm{L}_2,$ say

3. $\max_{\mathrm{H}_{03}} \mathrm{L} = \left(\frac{n\sqrt{(1-\rho_0^2)}}{\pi\eta_0}\right)^n \exp(-n)\mathrm{Q}^{-n}(\widehat{\mu}_{\mathrm{H}_{03}}) = \mathrm{L}_3,$ say

4. $\max_{\mathrm{H}_{04}} \mathrm{L} = \left(\frac{n\eta_0\sqrt{(1-\rho_0^2)}}{\pi}\right)^n \exp(-n)\mathrm{Q}^{-n}(\widehat{\mu}_{\mathrm{H}_{04}}) = \mathrm{L}_4,$ say

where L is likelihood function and $\mathrm{Q}(\widehat{\mu}) = \mathrm{Q}_1(\widehat{\mu}) + \mathrm{Q}_2$.

$$\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{01}}) = n\left[(\overline{x} - \widehat{\mu}_{\mathrm{H}_{01}})^2 - 2\frac{\rho_0}{\eta_0}(\overline{x} - \widehat{\mu}_{\mathrm{H}_{01}})(\overline{y} - \epsilon_0 - \widehat{\mu}_{\mathrm{H}_{01}}) + \frac{1}{\eta_0^2}(\overline{y} - \epsilon_0 - \widehat{\mu}_{\mathrm{H}_{01}})^2\right]$$

where $\widehat{\mu}_{\mathrm{H}_{01}} = a\overline{x} + (1-a)(\overline{y} - \epsilon_0)$, $a = \frac{\eta_0(\eta_0 - \rho_0)}{1+\eta_0^2 - 2\rho_0\eta_0}$.
Let, $K_1 = n(\overline{x} - \overline{y} + \epsilon_0)^2\left((1-a)^2 + 2\frac{\rho_0}{\eta_0}(1-a)a + \frac{a^2}{\eta_0^2}\right) > 0$ when $\eta_0 > 1$.
Therefore we can write $\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{01}}) = K_1,\ \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{02}}) = \eta_0^2 K_1,\ \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{03}}) = \frac{(\overline{x}-\overline{y}-\epsilon_0)^2}{(\overline{x}-\overline{y}+\epsilon_0)^2}K_1,$
$\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{04}}) = \frac{(\overline{x}-\overline{y}-\epsilon_0)^2}{(\overline{x}-\overline{y}+\epsilon_0)^2}\eta_0^2 K_1$.

We start with an assumption: $\epsilon_0 > 0, \quad \eta_0 > 1$.

Observation 1:

1. If $\overline{x} > \overline{y}$ then $\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{01}}) > \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{03}})$ and $\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{02}}) > \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{04}})$;
2. If $\overline{x} < \overline{y}$ then $\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{01}}) < \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{03}})$ and $\mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{02}}) < \mathrm{Q}_1(\widehat{\mu}_{\mathrm{H}_{04}})$.

Let $\mathrm{T}_1 = S_{xx} - 2.\frac{\rho_0}{\eta_0}S_{xy} + \frac{1}{\eta_0^2}S_{yy}$ and $\mathrm{T}_2 = \frac{1}{\eta_0^2}S_{xx} - 2.\frac{\rho_0}{\eta_0}S_{xy} + S_{yy}$.
Now $\mathrm{T}_2 - \mathrm{T}_1 = (1 - \frac{1}{\eta_0^2})(S_{yy} - S_{xx})$

Observation 2:

1. If $S_{yy} > S_{xx}$ and $\eta_0 > 1$ then $\mathrm{T}_2 > \mathrm{T}_1$;
2. If $S_{yy} < S_{xx}$ and $\eta_0 > 1$ then $\mathrm{T}_2 < \mathrm{T}_1$.

We observe that $\mathrm{Q}_{2,\mathrm{H}_{01}} = \mathrm{Q}_{2,\mathrm{H}_{03}} = \mathrm{T}_1$ and $\mathrm{Q}_{2,\mathrm{H}_{02}} = \mathrm{Q}_{2,\mathrm{H}_{04}} = \eta_0^2\mathrm{T}_2$.

Observation 3:

1. If $\overline{x} > \overline{y}$, then from observation 1, we conclude that $\mathrm{L}_1 < \mathrm{L}_3$ and $\mathrm{L}_2 < \mathrm{L}_4$;
2. If $\overline{x} < \overline{y}$, then from observation 1, we conclude that $\mathrm{L}_1 > \mathrm{L}_3$ and $\mathrm{L}_2 > \mathrm{L}_4$.

Now consider

$$\frac{\mathrm{L}_3}{\mathrm{L}_4} = \left(\frac{K_1^* + \mathrm{T}_2}{K_1^* + \mathrm{T}_1}\right)^n = \left(1 + \frac{\mathrm{T}_2 - \mathrm{T}_1}{K_1^* + T_1}\right)^n,$$

where $K_1^* = K_1 \frac{(\overline{x} - \overline{y} - \epsilon_0)^2}{(\overline{x} - \overline{y} + \epsilon_0)^2}$.

Observation 4:

1. If $S_{yy} > S_{xx}$ and $\eta_0 > 1$ then $\mathrm{L}_3 > \mathrm{L}_4$;
2. If $S_{yy} < S_{xx}$ and $\eta_0 > 1$ then $\mathrm{L}_3 < \mathrm{L}_4$.

Statement 1:

1. When $D1$: $\overline{x} > \overline{y}$; $S_x > S_y$ holds, and $\epsilon_0 > 0,\ \eta_0 > 1$ from Observation 3 and Observation 4, we conclude that $\mathrm{L}_4 > \mathrm{L}_1,\ \mathrm{L}_2\ , \mathrm{L}_3$;
2. When $D2$: $\overline{x} > \overline{y}$; $S_x < S_y$, and $\epsilon_0 > 0,\ \eta_0 > 1$ from Observation 3 and Observation 4, we conclude that $\mathrm{L}_3 > \mathrm{L}_1,\ \mathrm{L}_2\ , \mathrm{L}_4$.

Now consider

$$\frac{\mathrm{L}_1}{\mathrm{L}_2} = \left(\frac{K_1 + \mathrm{T}_2}{K_1 + \mathrm{T}_1}\right)^n = \left(1 + \frac{\mathrm{T}_2 - \mathrm{T}_1}{K_1 + T_1}\right)^n.$$

Observation 5:

1. If $S_{yy} > S_{xx}$ and $\eta_0 > 1$ then $\mathrm{L}_1 > \mathrm{L}_2$;
2. If $S_{yy} < S_{xx}$ and $\eta_0 > 1$ then $\mathrm{L}_1 < \mathrm{L}_2$.

Statement 2:

1. When $D3$: $\overline{x} < \overline{y}$; $S_x > S_y$, and $\epsilon_0 > 0,\ \eta_0 > 1$ from Observation 3 and Observation 5, we conclude that $\mathrm{L}_2 > \mathrm{L}_1,\ \mathrm{L}_3\ , \mathrm{L}_4$;
2. When $D4$: $\overline{x} < \overline{y}$; $S_x < S_y$, and $\epsilon_0 > 0,\ \eta_0 > 1$ from Observation 3 and Observation 5, we conclude that $\mathrm{L}_1 > \mathrm{L}_2,\ \mathrm{L}_3\ , \mathrm{L}_4$.

Conclusion: Statements 1 and 2 can be made without any further condition on the 4 statistics i.e. $\overline{x},\ \overline{y},\ S_x,\ S_y$.