İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 7, No. 1, January 2014, pp. 1–14
ISSN 1300-4077|14|1|1|14

İSTATİSTİK

A NOTE ON CONFIDENCE REGIONS BASED ON THE BIVARIATE CHEBYSHEV INEQUALITY. APPLICATIONS TO ORDER STATISTICS AND DATA SETS.

Jorge NAVARRO
Facultad de Matematicas, Universidad de Murcia, 30100 Murcia, Spain

Abstract: Chebyshev's inequality was recently extended to the multivariate case. In this paper this new inequality is used to obtain distribution-free confidence regions for an arbitrary bivariate random vector (X,Y). The regions depend on the means, the variances and the (Pearson) correlation coefficient. The theoretical method is illustrated by computing the confidence regions for two order statistics obtained from a sample of iid random variables or obtained from a sequence of dependent components. They are also computed for an arbitrary bivariate data set (with or without groups) by obtaining plots similar to univariate box plots.

Key words: Chebyshev (Tchebychev) inequality, Mahalanobis distance, Principal components, Ellipsoid, Order statistics, Bivariate box plots.
History: Submitted: 15 August 2014; Accepted: 9 September 2014

---

1. Introduction Chebyshev's inequality is a basic tool for random variables which provides a lower bound for the percentage of the population in a given distance with respect to the population mean when the variance is known. There are several extensions of this inequality to the multivariate case (see, e.g., [3, 5] and the references therein). The extension obtained recently by Chen [3] provided the following multivariate inequality

$$\Pr((\mathbf{X}-\mu)'V^{-1}(\mathbf{X}-\mu)\geq\varepsilon)\leq\frac{k}{\varepsilon} \tag{1.1}$$

valid for all $\varepsilon>0$ and for all random vectors $\mathbf{X}=(X_1,\ldots,X_k)'$ (where w' denotes the transpose of w) with finite mean vector $\mu=E(\mathbf{X})$ and positive definite covariance matrix $V=Cov(\mathbf{X})=E((\mathbf{X}-\mu)(\mathbf{X}-\mu)')$. This inequality can also be written as

$$\Pr((\mathbf{X}-\mu)'V^{-1}(\mathbf{X}-\mu)<\varepsilon)\geq 1-\frac{k}{\varepsilon} \tag{1.2}$$

for all $\varepsilon>0$ or as

$$\Pr(d_V(\mathbf{X},\mu)<\delta)\geq 1-\frac{k}{\delta^2} \tag{1.3}$$

for all $\delta>0$, where $d_V(\mathbf{X},\mu)=\sqrt{(\mathbf{X}-\mu)'V^{-1}(\mathbf{X}-\mu)}$ is the Mahalanobis distance associated to V between $\mathbf{X}$ and μ. Therefore (1.3) gives a lower bound for the probability in the *concentration ellipsoid* $E_\delta=\{\mathbf{x}\in\mathbb{R}^k: d_V(\mathbf{x},\mu)<\delta\}$.

Navarro [7] proved that the bounds in (1.1) are sharp, that is, they are the best possible bounds for these probabilities when only μ and V are known. A simple proof of (1.1) was obtained in [6] where the case of a singular covariance matrix is also considered by using the principal components associated to X. Some extensions of (1.1) to Hilbert-space-valued and Banach-space-valued random elements were given in [9] and [10], respectively. Additional (related) bounds are obtained in [2].

In this paper the inequality in (1.1) is used to obtain confidence regions for an arbitrary bivariate random vector (X,Y) (Section 2). The regions obtained depend on the means, the variances and the (Pearson) correlation coefficient. These confidence regions are computed in Section 3 for two order statistics obtained from a sample of iid random variables or obtained from a sequence of dependent random variables by using the expressions for the correlation coefficients given in [8]. They are also computed in Section 4 for an arbitrary bivariate data set (with or without groups) by obtaining plots similar to univariate box plots. Specifically, we consider the famous (Fisher's or Anderson's) iris data set.

2. Main result The following theorem proves that the bound in (1.1) can be used to obtain a confidence region for an arbitrary bivariate random vector (X,Y).

THEOREM 1. *Let $(X,Y)'$ be a random vector whose random variables have finite means $E(X)=\mu_X$ and $E(Y)=\mu_Y$, finite positive variances $Var(X)=\sigma_X^2>0$ and $Var(Y)=\sigma_Y^2>0$ and correlation coefficient $\rho=Cor(X,Y)\in(-1,1)$. Then*

$$\Pr((X^*-Y^*)^2+2(1-\rho)X^*Y^*<\delta)\geq 1-2\frac{1-\rho^2}{\delta} \tag{2.1}$$

for all $\delta>0$, where $X^*=(X-\mu_X)/\sigma_X$ and $Y^*=(X-\mu_Y)/\sigma_Y$.

T he covariance matrix of the random vector (X^*,Y^*) is

$$V=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$$

with $\rho=Cov(X^*,Y^*)=Cor(X,Y)\in(-1,1)$. Then

$$V^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix}$$

and, from (1.1), we obtain

$$\Pr\left(\frac{1}{1-\rho^2}(X^*,Y^*)\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix}\begin{pmatrix}X^*\\ Y^*\end{pmatrix}<\varepsilon\right)\geq 1-\frac{2}{\varepsilon},$$

for all $\varepsilon>0$, which gives

$$\Pr\left((X^*)^2+(Y^*)^2-2\rho X^*Y^*<\varepsilon(1-\rho^2)\right)\geq 1-\frac{2}{\varepsilon}$$

for all $\varepsilon>0$. Hence, by taking $\delta=\varepsilon(1-\rho^2)$, we obtain

$$\Pr\left((X^*)^2+(Y^*)^2+2(1-\rho-1)X^*Y^*<\delta\right)\geq 1-2\frac{1-\rho^2}{\delta},$$

that is,

$$\Pr\left((X^*-Y^*)^2+2(1-\rho)X^*Y^*<\delta\right)\geq 1-2\frac{1-\rho^2}{\delta}$$

for all $\delta>0$.

REMARK 1. Note that if $\rho=\pm 1$, then a similar expression can be obtained by using the univariate Chebyshev inequality. In [6] it is proved that if V is positive definite and T is an ortogonal

matrix (i.e. $T'T = TT' = I$) such that $T'VT = D$, where D is a diagonal matrix, then $V^{-1} = TD^{-1}T'$ and (1.2) can also be written as

$$\begin{aligned} p_\varepsilon &= \Pr((\mathbf{X}-\mu)'TD^{-1}T'(\mathbf{X}-\mu) < \varepsilon) \\ &= \Pr([D^{-1/2}T'(\mathbf{X}-\mu)]'[D^{-1/2}T'(\mathbf{X}-\mu)] < \varepsilon) \\ &= \Pr(\mathbf{Z}'\mathbf{Z} < \varepsilon) \\ &= \Pr(Z_1^2 + \cdots + Z_k^2 < \varepsilon) \geq 1 - \frac{k}{\varepsilon}, \end{aligned} \tag{2.2}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_k)' = D^{-1/2}T'(\mathbf{X}-\mu)$ are the standardized principal components. Hence (2.1) can also be obtained by computing the eigenvalues and eigenvectors of

$$V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The eigenvalues are $1+\rho$ and $1-\rho$ and the respective eigenvectors are $(1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$. Therefore $Z_1 = (X^* + Y^*)/\sqrt{2(1+\rho)}$ and $Z_2 = (X^* - Y^*)/\sqrt{2(1-\rho)}$ and, from (2.2), we obtain

$$\Pr\left(\frac{(X^*+Y^*)^2}{2(1+\rho)} + \frac{(X^*-Y^*)^2}{2(1-\rho)} < \varepsilon\right) \geq 1 - \frac{2}{\varepsilon} \tag{2.3}$$

for all $\varepsilon > 0$. A straightforward calculation shows that this expression is equivalent to (2.1). As we have mention in the introduction the regions determined by (2.1) (or by (2.3)) for X^* and Y^* correspond to ellipses with the main axes determined by the principal components obtained from the correlation matriz. They can be transformed into circles in terms of Z_1 and Z_2 by using (2.3) or into different ellipses in terms of the original random variables X and Y. Of course, if $\rho = 0$ and $\sigma_X = \sigma_Y$, then all these regions are circles. If (X, Y) have a bivariate normal distribution, the regions are determined by the level curves of the respective density function. Let us see a simple example.

Example 1. Let (X, Y) by an arbitrary random vector whose random variables have means $E(X) = E(Y) = 1$, variances $Var(X) = Var(Y) = 1$ and correlation coefficient $\rho = Cor(X, Y) = 0.9$. Then (2.1) gives

$$\Pr(5(X-Y)^2 + (X-1)(Y-1) < 5\delta) \geq 1 - 2\frac{0.19}{\delta},$$

that is,

$$\Pr(5X^2 - 9XY + 5Y^2 - X - Y + 1 < \varepsilon) \geq 1 - \frac{1.9}{\varepsilon}$$

for all $\varepsilon > 1.9$. The distribution-free confidence regions for $\varepsilon = 3, 4, 5, 10$ can be seen in Figure 1 containing at least the 36.6666%, 52.5%, 62% and the 81% of the values of (X, Y) (for any bivariate distribution). These regions can be simplified if we have some additional information about X and Y (e.g. that they are positive).

Remark 2. In the multivariate case (when $k > 2$), we can obtain similar plots for the standardized principal components $\mathbf{Z} = (Z_1, \ldots, Z_k)' = D^{-1/2}T'(\mathbf{X}-\mu)$ used in (2.2). For example, as $(Z_1, Z_2)'$ has mean $(0, 0)'$ and covariance matrix $Cov(Z_1, Z_2) = I_2$ (the identity matrix of dimension 2), then

$$\Pr(Z_1^2 + Z_2^2 < \varepsilon) \geq 1 - \frac{2}{\varepsilon}. \tag{2.4}$$

The same holds for the other pairs of principal components. A similar expression can be obtained for the usual principal components $\mathbf{Y} = (Y_1, \ldots, Y_k)' = T'\mathbf{X}$, that is, for $Y_i = E(Y_i) + \sqrt{\lambda_i}Z_i$, $i = 1, \ldots, k$. Thus, from (2.4), we obtain

$$\Pr\left(\frac{(Y_1 - E(Y_1))^2}{\lambda_1} + \frac{(Y_2 - E(Y_2))^2}{\lambda_2} < \varepsilon\right) \geq 1 - \frac{2}{\varepsilon}. \tag{2.5}$$

Let us see an example.

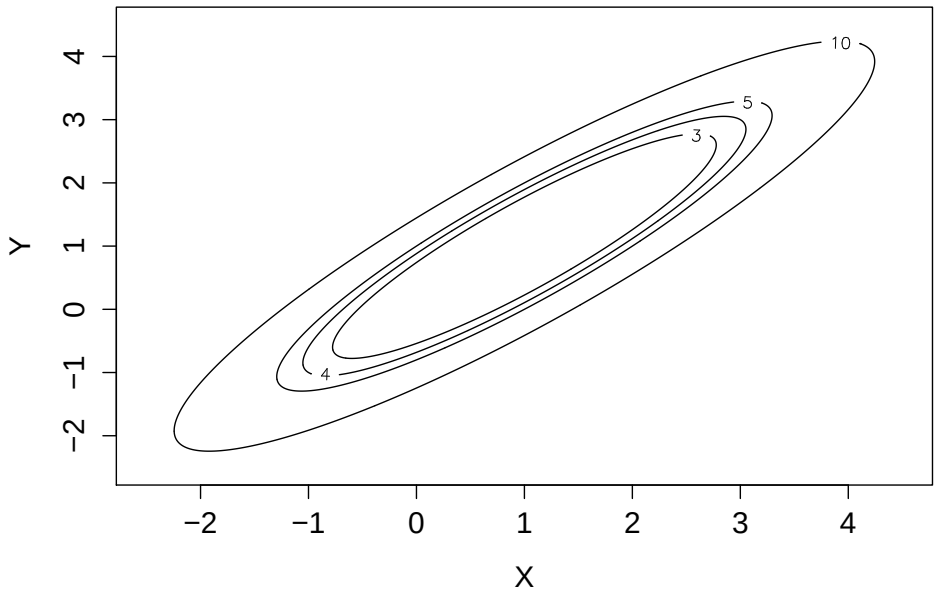

FIGURE 1. Confidence regions for the random vector in Example 1 for $\varepsilon = 3, 4, 5, 10$ containing at least the 36.6666%, 52.5%, 62% and the 81% of the values of (X, Y).

EXAMPLE 2. Let (X_1, X_2, X_3) by an arbitrary random vector with means $E(X_i) = 1$, variances $Var(X_i) = 1$ and correlation $\rho_{i,j} = 0.9$ for $i, j = 1, 2, 3$, $i \neq j$. Then the two greatest eigenvalues are $\lambda_1 = 1 + 2\rho = 2.8$, $\lambda_2 = 0.1$ and the (some) associated eigenvectors are $t_1 = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})'$ and $t_2 = (1/\sqrt{2}, -1/\sqrt{2}, 0)'$. Then, from (2.4), we have

$$\Pr(Z_1^2 + Z_2^2 < \varepsilon) \geq 1 - \frac{2}{\varepsilon}.$$

for $Z_1 = (X_1 + X_2 + X_3 - 3)/\sqrt{8.4}$ and $Z_2 = (X_1 - X_2)/\sqrt{0.2}$. Hence

$$\Pr\left(\frac{(X_1 + X_2 + X_3 - 3)^2}{8.4} + \frac{(X_1 - X_2)^2}{0.2} < \varepsilon\right) \geq 1 - \frac{2}{\varepsilon}.$$

Analogously, from (2.5), we obtain

$$\Pr\left(\frac{(Y_1 - 3)^2}{2.8} + \frac{Y_2^2}{0.1} < \varepsilon\right) \geq 1 - \frac{2}{\varepsilon},$$

where $Y_1 = (X_1 + X_2 + X_3)/\sqrt{3}$ and $Y_2 = (X_1 - X_2)/\sqrt{2}$. Thus, the distribution-free confidence regions for Y_1 and Y_2 and $\varepsilon = 4, 8$ can be seen in Figure 2 containing at least the 50% and the 75%, respectively, of the values of (Y_1, Y_2) (for any joint distribution).

3. Applications to order statistics Let us consider now the order statistics (ordered data) $X_{1:k}, \ldots, X_{k:k}$ obtained from the random sequence $X_1, \ldots, X_k$. This sequence can be the usual sample of independent and identically distributed (iid) random variables or they can be any random vector $(X_1, \ldots, X_k)$ (including a dependence structure, groups or outliers). They can also be seen in the context of the reliability theory as the lifetimes of j-out-of-k systems (i.e. systems which work when at least j of their k components work).

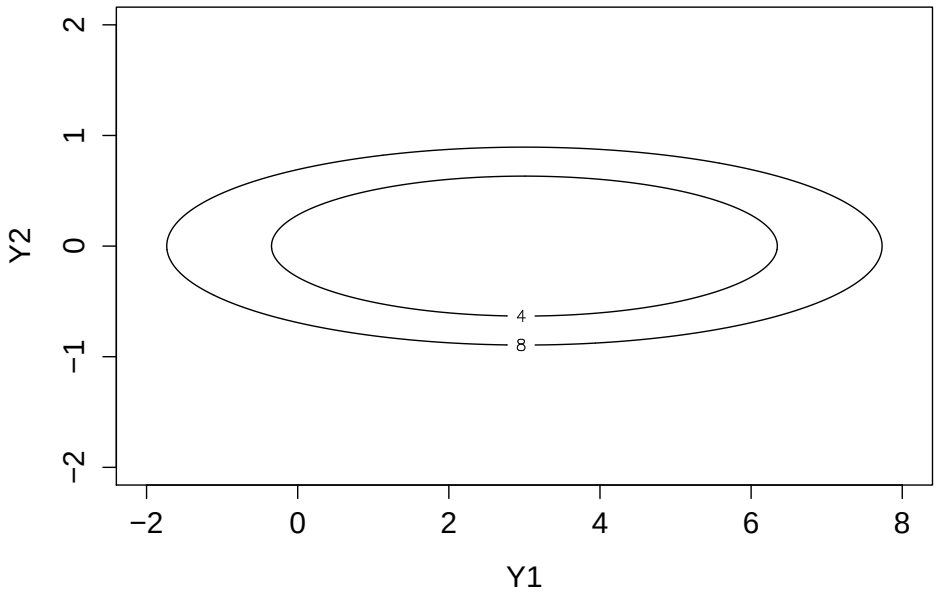

FIGURE 2. Confidence regions for the principal components of the random vector in Example 5 for $\varepsilon = 4, 8$ containing at least the 50% and the 75% of the values of (Y_1, Y_2).

In all these cases, a procedure to compute the respective correlation coefficients is given in [8]. Thus, for example, in the general case, for $k = 2$, from (2.2) in [8], we have

$$\rho_{1,2:2} = Cor(X_{1:2}, X_{2:2}) = \rho \frac{\sigma_1 \sigma_2}{\sigma_{1:2}\sigma_{1:2}} + \frac{(\mu_1 - \mu_{1:2})(\mu_2 - \mu_{1:2})}{\sigma_{1:2}\sigma_{1:2}},$$

where $\mu_i = E(X_i)$, $\mu_{i:2} = E(X_{i:2})$, $\sigma_i^2 = Var(X_i)$, $\sigma_{i:2}^2 = Var(X_{i:2})$, for $i = 1, 2$, and $\rho = Cor(X_1, X_2)$. In particular, if X_1 and X_2 are identically distributed (id) then

$$\rho_{1,2:2} = Cor(X_{1:2}, X_{2:2}) = \rho \frac{\sigma^2}{\sigma_{1:2}\sigma_{1:2}} + \frac{(\mu - \mu_{1:2})^2}{\sigma_{1:2}\sigma_{1:2}},$$

where $\mu = E(X_i)$ and $\sigma^2 = Var(X_i)$ for $i = 1, 2$. Hence, in both cases, from Theorem 1, we have

$$\Pr((X_{2:2}^* - X_{1:2}^*)^2 + 2(1 - \rho_{1,2:2})X_{2:2}^* X_{1:2}^* < \delta) \geq 1 - 2\frac{1 - \rho_{1,2:2}^2}{\delta}, \tag{3.1}$$

where $X_{i:2}^* = (X_{i:2} - \mu_{i:2})/\sigma_{i:2}$, $i = 1, 2$. Notice that in this case we also have the relation $X_{2:2} \geq X_{1:2}$. By combining this expression with (3.1) we can obtain confidence regions for $(X_{1:2}, X_{2:2})$. Note that this can also be applied to the study of a two components parallel system (with lifetime $X_{2:2} = \max(X_1, X_2)$) and the time of the first component failure ($X_{1:2} = \min(X_1, X_2)$). Here we might also have some additional restrictions (e.g. $X_{1:2} \geq 0$). Let us see some examples.

EXAMPLE 3. If (X_1, X_2) has an exchangeable Pareto distribution with joint reliability function

$$\overline{F}(x, y) = \Pr(X_1 > x, X_2 > y) = (1 + \lambda x + \lambda y)^{-\theta}$$

for $x, y \geq 0$, where $\lambda > 0$ and $\theta > 2$, then $\mu = 1/(\lambda\theta - \lambda)$, $\sigma^2 = \mu^2/(1 - 2\rho)$ and $\rho = 1/\theta \in (0, 1/2)$. Moreover, $\mu_{1:2} = \mu/2$, $\mu_{2:2} = 3\mu/2$

$$\sigma_{1:2}^2 = \frac{\mu^2}{4(1 - 2\rho)},$$

$$\sigma^2_{2:2} = \frac{\mu^2(6+3\rho)}{4(1-2\rho)}$$

and

$$\rho_{1,2:2} = \frac{1+2\rho}{\sqrt{6+3\rho}}$$

(see Example 2.10 in [8]). For example, if $\lambda = 0.5$ and $\theta = 3$, then $\mu = 1$, $\rho = 1/3$, $\mu_{1:2} = 1/2$, $\mu_{2:2} = 3/2$, $\sigma_{1:2} = 0.8660254$, $\sigma_{2:2} = 2.291288$ and $\rho_{1,2:2} = 0.6299408$. Hence, from (3.1), we have

$$\Pr\left(\left(\frac{X_{2:2}-1.5}{2.291288} - \frac{X_{1:2}-0.5}{0.8660254}\right)^2 + 0.7401\frac{X_{2:2}-1.5}{2.291288}\frac{X_{1:2}-0.5}{0.8660254} < \delta\right) \geq 1 - \frac{1.2063}{\delta}$$

for all $\delta > 0$. The confidence regions for $(X_{1:2}, X_{2:2})$ obtained from this expression and $0 \leq X_{1:2} \leq X_{2:2}$ for $\delta = 2, 4, 6$ are plotted in Figure 3. They contain at least the 39.68254%, the 69.84127% and the 79.89418%, respectively, of the values of $(X_{1:2}, X_{2:2})$.

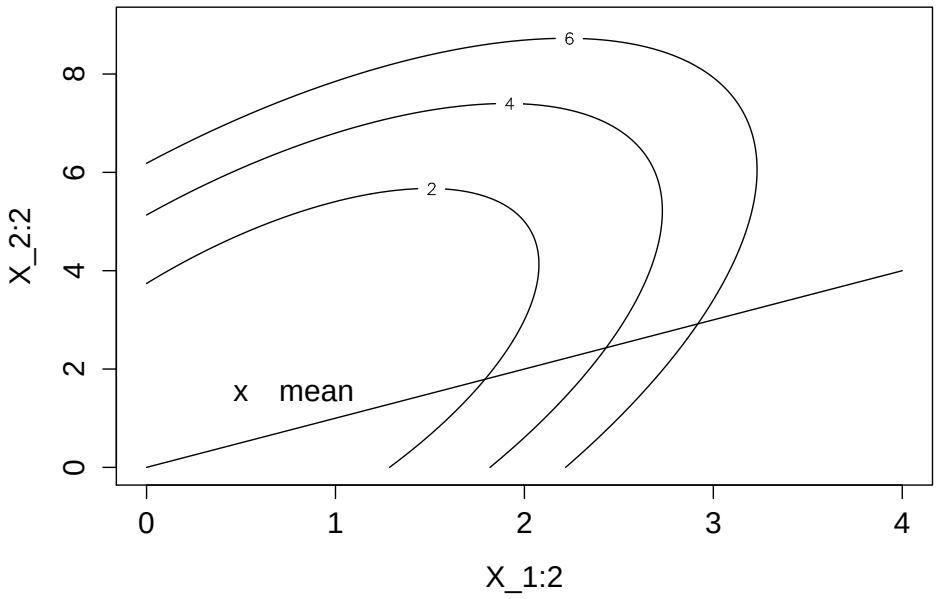

FIGURE 3. Confidence regions for the order statistics when (X_1, X_2) have a Pareto distribution (see Example 3) with $\lambda = 0.5$ and $\theta = 3$ for $\delta = 2, 4, 6$ containing at least the 39.68254%, the 69.84127% and the 79.89418% of the values of $(X_{1:2}, X_{2:2})$.

EXAMPLE 4. If (X_1, X_2) has an exchangeable normal distribution with common mean μ, common variance σ^2 and correlation $\rho \in (-1, 1)$, then (see Example 2.9 in [8])

$$\mu_{1:2} = \mu - \sigma\sqrt{\frac{1-\rho}{\pi}},$$

$$\mu_{2:2} = \mu + \sigma\sqrt{\frac{1-\rho}{\pi}},$$

$$\sigma_{1:2} = \sigma_{2:2} = \sigma\sqrt{1 - \frac{1-\rho}{\pi}}$$
$$\rho_{1,2:2} = \frac{1 + (\pi - 1)\rho}{\pi - 1 + \rho},$$

(see Example 2.9 in [8]). For example, if $\mu = 3$, $\sigma = 1$ and $\rho = 0.5$, then $\mu_{1:2} = 2.601058$, $\mu_{2:2} = 3.398942$, $\sigma_{1:2} = \sigma_{2:2} = 0.916976$ and $\rho_{1,2:2} = 0.7839196$. Hence, from (3.1), we have

$$\Pr\left(\left(\frac{X_{2:2} - 3.4}{0.916976} - \frac{X_{1:2} - 2.6}{0.916976}\right)^2 + 0.432\frac{X_{2:2} - 3.4}{0.916976}\frac{X_{1:2} - 2.6}{0.916976} < \delta\right) \geq 1 - \frac{0.77094}{\delta}$$

for all $\delta > 0$. The confidence regions for $(X_{1:2}, X_{2:2})$ obtained from this expression and $X_{1:2} \leq X_{2:2}$ for $\delta = 2, 3, 4$ are plotted in Figure 4. They contain at least the 61.453%, the 74.302% and the 80.7265%, respectively, of the values of $(X_{1:2}, X_{2:2})$.

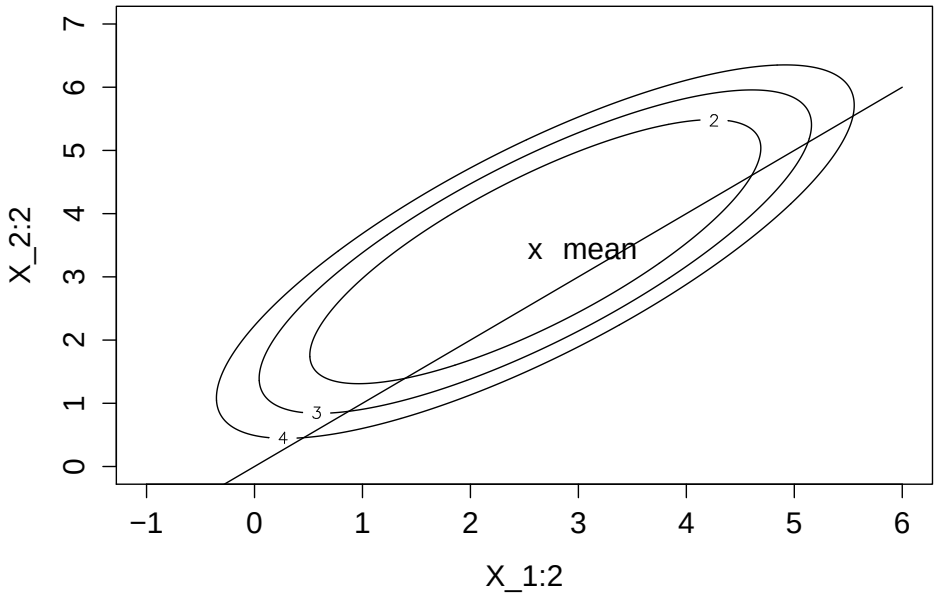

FIGURE 4. Confidence regions for the order statistics when (X_1, X_2) have an exchangeable Normal distribution (see Example 4) with $\mu = 3$, $\sigma = 1$ and $\rho = 0.5$ for $\delta = 2, 3, 4$ containing at least the 61.453%, the 74.302% and the 80.7265% of the values of $(X_{1:2}, X_{2:2})$.

Similar confidence regions can be obtained for the general order statistics $X_{i:k}$ and $X_{j:k}$ from Theorem 1 whenever we can compute their means, variances and correlation coefficients. If these order statistics come from a sample of iid random variables, then their means, variances and correlation coefficients can be obtained from the expressions given Theorem 3.3 of [8].

In particular, if the order statistics come from a sample of iid exponential distributions, then the correlation coefficients for $k \leq 6$ are given in Table 1 of [8]. Alternatively, in this case, if $E(X_i) = 1$ for $i = 1, \ldots, k$, then they can be computed from the following well known (see [1]) expressions

$$\mu_{i:k} = E(X_{i:k}) = \sum_{j=k-i+1}^{k} \frac{1}{j},$$

$$\sigma^2_{i:k} = Var(X_{i:k}) = \sum_{j=k-i+1}^{k} \frac{1}{j^2}$$

and

$$\rho_{i,j:k} = Cor(X_{i:k}, X_{j:k}) = \frac{\sigma_{i:k}}{\sigma_{j:k}}$$

for $1 \leq i < j \leq k$. Hence we can use Theorem 1 to obtain confidence regions for $(X_{i:k}, X_{j:k})$. For example, if $k = 3$, $i = 2$ and $j = 3$, then $\mu_{2:3} = 5/6$, $\mu_{3:3} = 11/6$, $\sigma_{2:3} = 0.6009252$, $\sigma_{3:3} = 1.166667$, and $\rho_{2,3:3} = 0.5150788$. Hence, from (2.1), we have

$$\Pr\left(\left(\frac{X_{3:3} - 11/6}{1.166667} - \frac{X_{2:3} - 5/6}{0.6009252}\right)^2 + 0.9698\frac{X_{3:3} - 11/6}{1.166667}\frac{X_{2:3} - 5/6}{0.6009252} < \delta\right) \geq 1 - \frac{1.4694}{\delta}$$

for all $\delta > 0$. The confidence regions for $(X_{2:3}, X_{3:3})$ obtained from this expression and $0 \leq X_{2:3} \leq X_{3:3}$ for $\delta = 4, 6, 8$ are plotted in Figure 5. They contain at least the 63.2653%, the 75.5102% and the 81.6326%, respectively, of the values of $(X_{2:3}, X_{3:3})$.

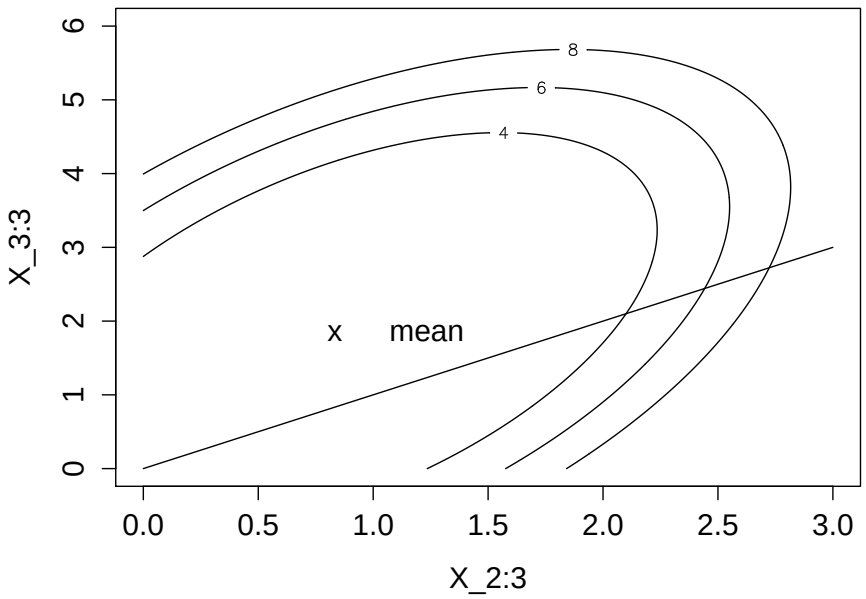

FIGURE 5. Confidence regions for the order statistics when X_1, X_2, X_3 are iid with an exponential distribution with mean 1 for $\delta = 2, 3, 4$ containing at least 63.2653%, the 75.5102% and the 81.6326% of the values of $(X_{2:3}, X_{3:3})$.

Analogously, we can consider the random vector $(X_{1:k}, \ldots, X_{k:k})'$ and to obtain confidence regions for it from (1.2). Thus, we obtain

$$\Pr(\mathbf{X}'_{OS} R^{-1} \mathbf{X}_{OS}) < \varepsilon) \geq 1 - \frac{k}{\varepsilon}$$

for all $\varepsilon > 0$, where $\mathbf{X}_{OS} = (X^*_{1:k}, \ldots, X^*_{k:k})$ and $X^*_{i:k} = (X_{i:k} - \mu_{i:k})/\sigma_{i:k}$ for $i = 1, \ldots, k$ and $R = Cor(X_{1:k}, \ldots, X_{k:k})'$. For example, in the present case of exponential distributions with common mean 1, if $k = 3$ we obtain

$$R = \begin{pmatrix} 1 & 0.5547002 & 0.2857143 \\ 0.5547002 & 1 & 0.5150788 \\ 0.2857143 & 0.5150788 & 1 \end{pmatrix}$$

(see Table 1 in [8]) and the confidence region

$$R_\varepsilon = \{(x, y, z) : 1.444x^2 - 1.602xy + 1.805y^2 - 1.402yz + 1.361z^2 < \varepsilon\}$$

containing $\mathbf{X}_{OS} = (X^*_{1:3}, X^*_{2:3}, X^*_{3:3})'$ with a probability greater than $1 - k/\varepsilon$. A similar confidence region can be obtained for $(X_{1:3}, X_{2:3}, X_{3:3})'$.

In a similar way, we can use the expressions (2.2) or (2.4) for the principal components obtained from the order statistics. For example, with the last one, from the matrix R above, we obtain

$$\Pr\left(\frac{Y_1^2}{1.9129431} + \frac{Y_2^2}{0.77153779} < \varepsilon\right) \geq 1 - \frac{2}{\varepsilon} \tag{3.2}$$

for all $\varepsilon > 0$, where

$$Y_1 = 0.5548133X^*_{1:3} + 0.6382230X^*_{2:3} + 0.5337169X^*_{3:3}$$

and

$$Y_2 = 0.66914423X^*_{1:3} + 0.03890251X^*_{2:3} - 0.7421136X^*_{3:3}.$$

The confidence regions obtained from this expression for $\varepsilon = 4, 6, 8$ are plotted in Figure 6. They contain at least the 50%, the 66.6667% and the 75%, respectively, of the values of $(X_{1:3}, X_{2:3}, X_{3:3})$.

4. Applications to samples The confidence regions obtained from Theorem 1 can be used to obtain bivariate plots similar to (univariate) box plots. Thus if we have a sample $O_i = (X_i, Y_i)'$, $i = 1, \ldots, n$ from $(X, Y)'$ (i.e. iid random vectors equal in law to $(X, Y)'$) or just a collection of n pairs of data (they can be dependent and with different distributions), we can consider the empirical (discrete) distribution associated to the data set which choose the data O_i with probability $1/n$. The mean of this discrete distribution is of course

$$\overline{O} = \frac{1}{n}\sum_{i=1}^{n} O_i = (\overline{X}, \overline{Y})$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Analogously, its covariance matrix is

$$\widehat{V} = \frac{1}{n}\sum_{m=1}^{n}(O_m - \overline{O})(O_m - \overline{O})' = (\widehat{V}_{i,j}),$$

where

$$\widehat{V}_{1,1} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

$$\widehat{V}_{2,2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

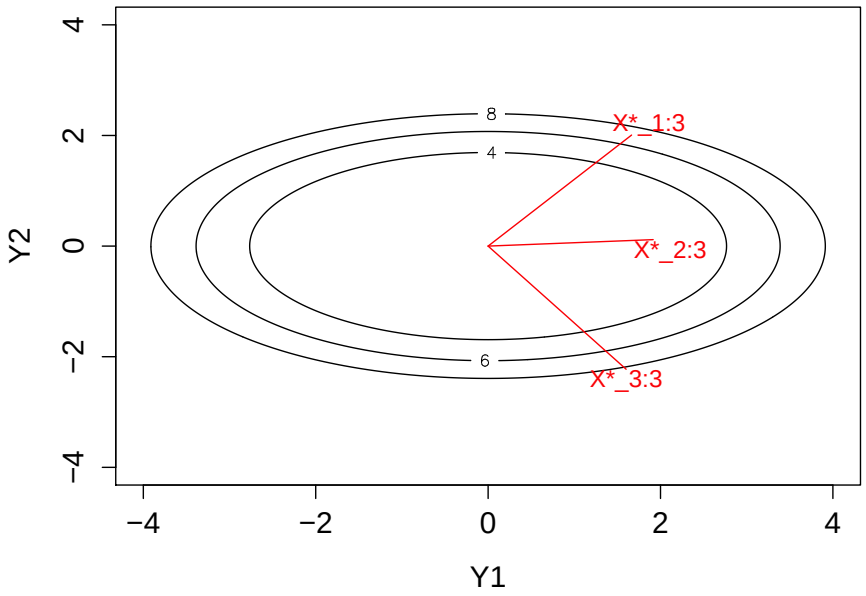

FIGURE 6. Confidence regions obtained from (3.2) for the principal components of the order statistics when X_1, X_2, X_3 are iid with an exponential distribution with mean 1 for $\varepsilon = 4, 6, 8$ containing at least the 50%, the 66.6667% and the 75% of the values of $(X_{1:3}, X_{2:3}, X_{3:3})$.

and

$$\widehat{V}_{1,2} = \widehat{V}_{2,1} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

. Hence the correlation coefficient is

$$r = \frac{\widehat{V}_{1,2}}{\sqrt{\widehat{V}_{1,1}\widehat{V}_{2,2}}}.$$

Then, from Theorem 1, if $-1 < r < 1$, we have

$$\Pr((X_I^* - Y_I^*)^2 + 2(1-r)X_I^* Y_I^* < \delta) \geq 1 - 2\frac{1-r^2}{\delta} \tag{4.1}$$

for all $\delta > 0$, where $X_I^* = (X_I - \overline{X})/\sqrt{\widehat{V}_{1,1}}$, $Y_I^* = (Y_I - \overline{Y})/\sqrt{\widehat{V}_{2,2}}$ and $I = i$ with probability $1/n$ for $i = 1, 2, \ldots, n$, that is, (X_I, Y_I) is a randomly chosen data in the data set $\{O_i = (X_i, Y_i)', \quad i = 1, \ldots, n\}$.

Then, by taking $\delta = 4(1-r^2)$ in (4.1), the confidence (elliptical) region

$$R_1 = \{(x, y) \in \mathbb{R}^2 : (x^* - y^*)^2 + 2(1-r)x^* y^* < 4(1-r^2)\},$$

where $x^* = (x - \overline{X})/\sqrt{V_{1,1}}$, $y^* = (y - \overline{Y})/\sqrt{V_{2,2}}$, contains (for sure) at least the 50% of the data (in this data set).

Analogously, by taking $\delta = 8(1-r^2)$ in (4.1), the confidence (elliptical) region

$$R_2 = \{(x, y) \in \mathbb{R}^2 : (x^* - y^*)^2 + 2(1-r)x^* y^* < 8(1-r^2)\},$$

contains (for sure) at least the 75% of the data and the region

$$R_3 = \{(x,y) \in \mathbb{R}^2 : (x^* - y^*)^2 + 2(1-r)x^*y^* \geq 8(1-r^2)\},$$

contains (for sure) at most the 25% of the data.

These regions can be used to obtain bivariate plots similar to (univariate) box plots containing (for sure) more than the 50% and 75% of the data (R_1 and R_2) and less than the 25% of the data (R_3), respectively. Note that n can be as big as we want and hence this procedure can be applied to (very) big data sets even if they come from different populations, contain outliers and/or are dependent (e.g. time-dependent). Obviously, the regions will be more accurate if we really have a sample from a given population (distribution) or if they are computed in each group (if there are different groups). Similar regions can be obtained from (2.4) and (2.5) (or from (4.1) for the scores in a principal components analysis. Let us see a simple example. A similar procedure can be applied to the canonical projections in a discriminant analysis.

Example 5. Let us consider the data set called "*iris*" included in the statistical program R. This famous (Fisher's or Anderson's) iris data set (see [4]) gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 iris flowers from each of 3 species of iris. The species are iris setosa, versicolor, and virginica. In this data set we consider the variables $X = Petal.Length$ and $Y = Petal.Width$. For these variables (without consider separate groups for each specie), we obtain $r = 0.9628654$ and the confidence regions R_1 and R_2 determined by

$$\left(\frac{x-3.758}{1.759404} - \frac{y-1.199333}{0.7596926}\right)^2 + 2(1-r)\frac{x-3.758}{1.759404}\frac{y-1.199333}{0.7596926} < 0.2915606$$

and

$$\left(\frac{x-3.758}{1.759404} - \frac{y-1.199333}{0.7596926}\right)^2 + 2(1-r)\frac{x-3.758}{1.759404}\frac{y-1.199333}{0.7596926} < 0.5831213,$$

respectively. The confidence regions and the data can be seen in Figure 7. Note that, of course, these regions contain more than the 50% and the 75% of the data (i.e. more than 75 and 113 data in this case) and the outside region R_3 contains less than the 25% of the data (in this case it contains only 2 data which is less than 37). Note that the data come from three different populations (species). Also note that if the data set is big, then we do not need to plot the data. In a similar way, we can compute the confidence regions for the data in each group obtaining the plot in Figure 8.

Analogously, we can consider the two first principal components Y_1 and Y_2 computed from the correlation matrix of the four variables in this data set. The two first eigenvalues are 2.91849782 and 0.91403047 and the principal components are obtained as

$$Y_1 = 0.5210659X_1^* - 0.2693474X_2^* + 0.5804131X_3^* + 0.5648565X_4^*$$

and

$$Y_2 = -0.37741762X_1^* - 0.92329566X_2^* - 0.02449161X_3^* - 0.06694199X_4^*,$$

where $X_i^* = (X_i - \overline{X}_i)/\sqrt{\widehat{V}_{i,i}}$, $i = 1,2,3,4$ are the standardized versions of the original variables $X_1 = sepal.length$ $X_2 = sepal.width$, $X_3 = petal.length$ and $X_4 = petal.width$. In this case, $\overline{Y}_1 = \overline{Y}_2 = 0$ and $r = 0$ and hence the confidence regions R_1 and R_2 are determined by

$$\frac{x^2}{2.91849782} + \frac{y^2}{0.91403047} < 4$$

and

$$\frac{x^2}{2.91849782} + \frac{y^2}{0.91403047} < 8,$$

respectively. These regions will contain (for sure) the 50% and 75% of the data scores for these principal components. These scores and the regions are plotted in Figure 9.

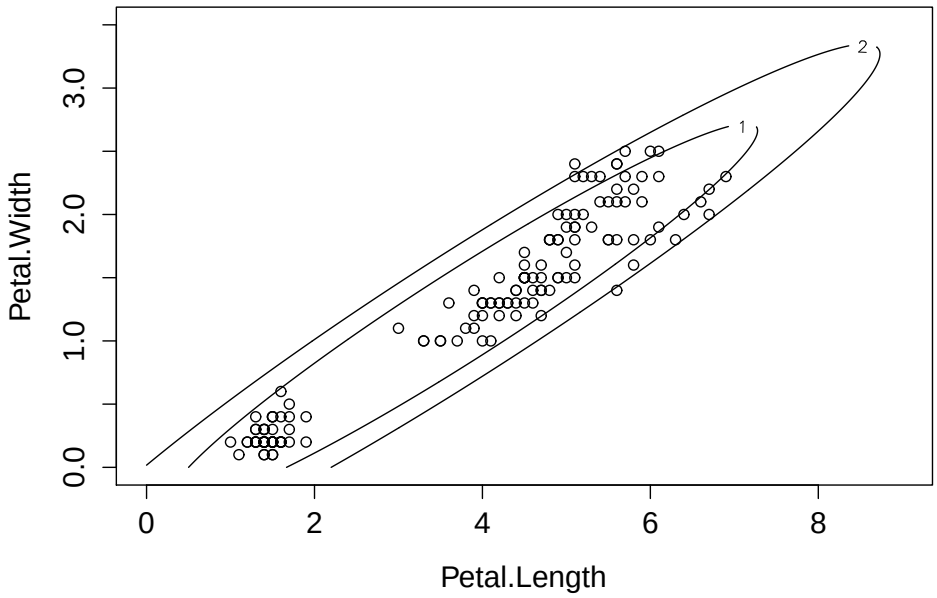

FIGURE 7. Confidence regions R_1 and R_2 for the variables $X = Petal.Length$ and $Y = Petal.Width$ included in the data set "iris" considered in Example 5 containing (for sure) at least the 50% and 75% of the data.

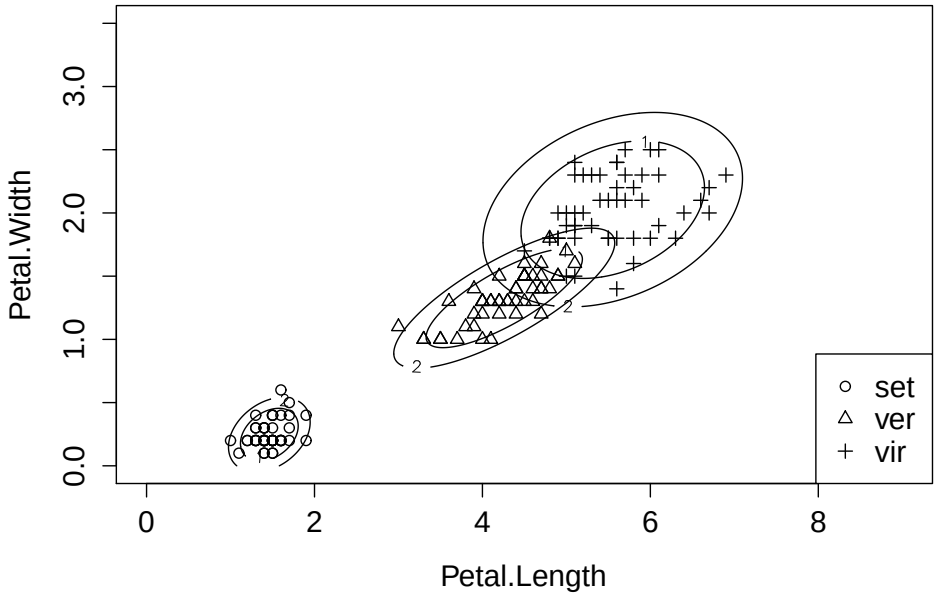

FIGURE 8. Confidence regions R_1 and R_2 for the variables $X = Petal.Length$ and $Y = Petal.Width$ included in the data set "iris" considered in Example 5 containing (for sure) at least the 50% and 75% of the data by species.

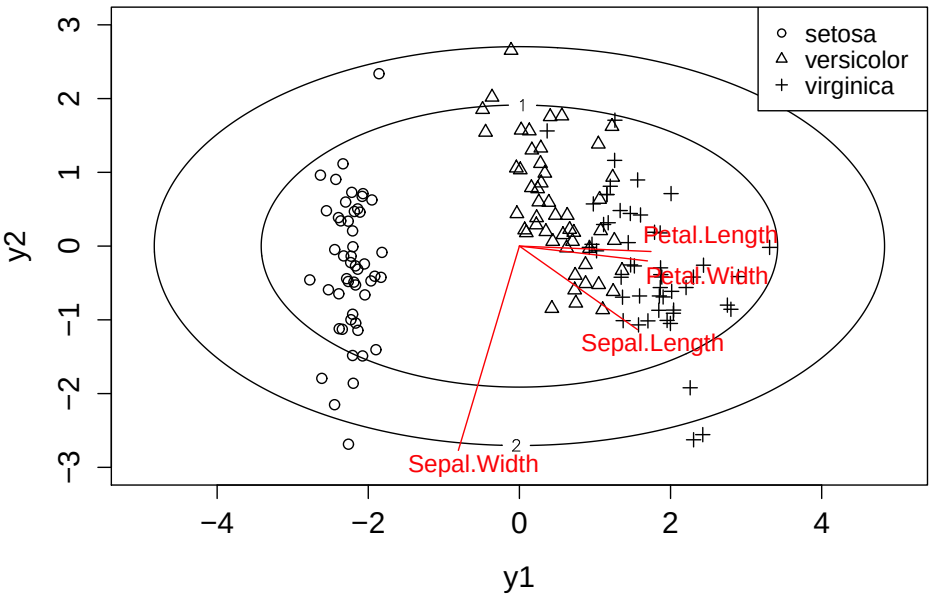

FIGURE 9. Confidence regions R_1 and R_2 for the scores in the two first principal components of the variables included in the data set "iris" considered in Example 5 containing (for sure) at least the 50% and 75% of the data scores.

Note that we can obtain regions of this type containing exactly the 50% or the 75% of the data by computing the median and the third quartile of Mahalanobis distances $d_{\widehat{V}}(O_i, \overline{O})$, $i = 1, \ldots, n$. In this case we do not need to use Theorem 1.

Finally, it is important to note that all the confidence regions obtained in this sections are only for the data in the data set. We may wonder if these confidence regions will contain a new data (obtained by a similar procedure). In this case we may use a standard cross validation technique to compute the confidence probabilities for this new data in each data set (i.e., we can compute the confidence region without using a given data in the data set and then study if this region contains this data).

5. Conclusions Theorem 1 gives a distribution free confidence region for (X, Y) based on the means, the variances and the correlation coefficient. As the bounds are sharp, this is the best confidence region that can be obtained without additional assumptions. This region can be applied to any pair of related random variables. In this paper we apply them theoretical models, principal components and the order statistics. We also apply them to data sets by using the sample measures obtaining bivariate plots similar to univariate box plots.

Acknowledgements This work is partially supported by Ministerio de Economa y Competitividad under grant MTM2012-34023-FEDER and Fundacin Sneca of C.A.R.M. under grant 08627/PI/08.

References

[1] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N.(2008). *A First Course in Order Statistics*. Classic ed., SIAM, Philadelphia, Pennsylvania.

[2] Budny, K. (2014). A generalization of Chebyshev's inequality for Hilbert-space-valued random elements. *Statistics & Probability Letters*, 88, 62-65.

[3] Chen, X. (2011). A new generalization of Chebyshev inequality for random vectors. ArXiv:0707.0805v2.

[4] Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179-188.

[5] Marshall, A.W. and Olkin, I. (1960). Multivariate Chebyshev inequalities. *The Annals of Mathematical Statistics*, 31, 1001-1014.

[6] Navarro, J. (2014). A very simple proof of the multivariate Chebyshev's inequality. DOI:10.1080/03610926.2013.873135.

[7] Navarro, J. (2014). Can the bounds in the multivariate Chebyshev inequality be attained?. *Statistics & Probability Letters*, 91, 1-5.

[8] Navarro, J. and Balakrishnan, N. (2010). Study of some measures of dependence between order statistics and systems. *Journal of Multivariate Analysis*, 101, 52-67.

[9] Prakasa Rao, B.L.S. (2010). Chebyshev's inequality for Hilbert-space-valued random elements. *Statistics & Probability Letters*, 80, 1039-1042.

[10] Zhou, L. and Hu, Z.C. (2012). Chebyshev's inequality for Banach-space-valued random elements. *Statistics & Probability Letters*, 82, 925-931.