1. Introduction

The correlation coefficient ρ is a very popular statistic that has wide applications in many fields, such as education, economy, chemistry, biology, medicine,..., etc. Suppose $\{X_{1,...}X_n\}$ and $\{Y_{1,...}Y_n\}$ are two independent samples on (X, Y), the Pearson Correlation Coefficient is definded as

$$r = \frac{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)\left(Y_j - \overline{Y}\right)}{\sqrt{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2 \sum_{j=1}^{n} \left(Y_j - \overline{Y}\right)^2}},$$

which is a sample version of the population correlation defined as

$$\rho = E\left[(x - E(X))(y - E(Y))\right] / \sigma_x \sigma_y,$$

where

$$\sigma_x^2 = E\left[x - E(X)\right]^2,$$

$$\sigma_y^2 = E\left[y - E(Y)\right]^2.$$

In this thesis, we will develop reliable confidence intervals for the difference of the correlation coefficients from two populations. We will base our proposals on the two corresponding Pearson correlation coefficients that are obtained from two independent samples from the respective populations. There are situations where the difference of two correlation coefficients is of primary interent. For example, to find a region of genes that are associated with a certain disease, it can be helpful to compare the correlations between genetic factors among the case group and the control group (see for example Zaykin et al. 2006). If we view the properties regarding the mean (or median) as the "first order" characteristic, then the properties regarding the correlations forms a sort of "second order" characteristic and it is reasonable to compare the second-order characteristic among two groups of subjects in addition to the first-order comparions. Although there exists plenty of statistical inference

procedures for the correlation coefficient in one-sample seetings and a few testing procedures in two-sample settings, see for example Muirhead (1982, chap.5), to our knowledge there are few procedures in the liferature developed for the confidence interval for correlation coefficients in two-sample settings.

We will develop three confidence intervals for the difference of two correlation coefficients based on two independent samples: the first one (confidence interval I) is derived from simple applications of the Law of Large Number and the Slutsky Theorem, the second one (confidence interval II) is derived according to a result from bivariate elliptical distribution, and the third one (confidence interval III) is obtained by applying the Fisher's z transformation(Fisher 1921) and an inverse transformation. Our simulation studies reveal that all the three confidence intervals described above have correct coverage probabilities (CPs) in most of the situations we considered, although the confidence interval I tends to have CPs higher than norminal values when both the two population correlations are large in magnitude, and the confidence interval III tends to have CPs lower than norminal values when one of the two population correlations is large in magnitude and the sample sizes are highly imbalanced among the two samples.

On the other hand, regarding the precision (length) of the confidence interval, the length of confidence interval III is almost always shorter than that of the confidence interval II, and the length of confidence interval II tends to be shorter than that of confidence interval I. Therefore, for practical applications, we would suggest using the confidence interval III when the two correlations are moderate in magnitude or when the two sample sizes are balenced, while in other situations the confidence interval II is a better choice.

This thesis is organized as follows. In Section 2 we present some preliminary results relevant to our problems. The proposed three confidence intervals are described in Section 3. Section 4 displays our simulation results, and Section 5 displays a real data analysis. We conclude this thesis by some further discussions in Section 6.