TESTING THE HYPOTHESIS OF MULTIVARIATE NORMALITY IN BAYESIAN APPROACHES TO SPEAKER ADAPTATION

Li-Wei WANG
Department of Electronic and Engineering, Tsinghua University, China
wanglw@thsp.ee.tsinghua.edu.cn

Zuo-Ying WANG
Department of Electronic and Engineering, Tsinghua University, China
wzy-dee@mail.tsinghua.edu.cn

ABSTRACT
Bayesian approaches to speaker adaptation are popular in Automatic Speech Recognition (ASR) systems. In most kinds of Bayesian adaptation, there are parameters whose prior distributions are assumed to be multivariate normal. This paper presents a methodology, which can test the hypothesis of multivariate normality. When applied to Maximum A Posteriori (MAP) adaptation, we found that the real prior distributions of the mean vectors are far from normal, which are always assumed in the MAP procedure. This result implies that better choice of the prior form may improve the adaptation result.

1. INTRODUCTION
In Bayesian approaches to speaker adaptation [3], [4], [6], the parameters of Hidden Markov Models (HMM) [5] are looked upon as random variables having prior distributions. A good choice of the form of prior distribution will lead to an accurate estimate of the parameters. Some MAP-type adaptation methods assume multivariate normal distributions to the prior form of mean vector of each Gaussian in the mixture density of the observations [4], [6]. However, there are few experiments on whether such an assumption is correct and how far the assumed form departures from the real prior form. The experiments in [7] tested the posterior distribution of logarithmic spectral energy and showed that it was strongly non-normal. The hypothesis test in [7] is only of use when the parameter being tested is a one-dimensional random variable. This paper is concerned with whether the multivariate normality assumption is correct to the multidimensional mean vectors. Since marginal distributions are univariate normal is not sufficient to a multivariate normal distribution, this method could not test the prior distributions of the mean vectors of the Gaussians.

In this work we develop a hypothesis test to decide whether the mean vectors are distributed according to multivariate normal distribution. Experimental results show that the real prior form is far from multivariate normality. This implies that better choice of the prior form may improve the adaptation result.

2. TESTS FOR UNIVARIATE NORMALITY
In the test for univariate normality, the null hypothesis \( H_0 \) is that the distribution of the random variable \( \eta \) is univariate normal. Given the significance level \( \alpha \) (usually chosen as 0.01 or 0.05), the probability of type I error is completely controlled by \( \alpha \), that is
\[
P(\alpha) = P\{\text{rejection of } H_0/ H_0 \text{ is true}\} = P\{\text{Decide the distribution of } \eta \text{ is not normal } | \eta \text{ is normal}\} = \alpha.
\]

Given the samples \( \{x_i, i = 1,2,\cdots, N\} \), the following two statistics can be used to test whether the distribution is univariate normal.

Skewness test
\[
\overline{S}_x = \frac{1}{nS^3} \sum_{i=1}^{n} (x_i - \overline{x})^3.
\]

Kurtosis test
\[
\overline{E}_x = \frac{1}{NS^4} \sum_{i=1}^{n} (x_i - \overline{x})^4 - 3.
\]

Where \( \overline{x} \) and \( S \) are sample mean and sample standard deviation [1]. The critical value of \( \overline{S}_x \) and \( \overline{E}_x \) are their quantile of order \( \alpha/2 \) respectively (tabulated in [1]). If the statistic is larger than the critical value, we decide that the distribution is non-normal while the probability of committing an error is \( \alpha \).

3. TESTS FOR MULTIVARIATE NORMALITY
In this section we develop a new method to test the multivariate normality of random vector. Similar to the previous section, the null hypothesis is that the distribution of the random vector is multivariate normal. Our method mainly based on the following theorem.

Theorem 1.
A necessary and sufficient condition that a $p$-component random vector $\xi = (\xi_1, \xi_2, \ldots, \xi_p)^T$ is distributed according to multivariate normal distribution is that all the linear combinations of $\xi_1, \xi_2, \ldots, \xi_p$ are distributed according to univariate normal distribution. That is, for any vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)^T$, the random variable

$$\eta = \lambda^T \xi = \sum_{i=1}^{p} \lambda_i \xi_i .$$

is univariate normal.

The proof of the theorem can be found in [2].

This theorem shows a relationship between the test of multivariate normality with the tests of univariate normality of the linear combinations. Inspired by the theorem, we construct $M$ linear combinations $\eta_1, \eta_2, \ldots, \eta_M$

$$\eta_m = \lambda^T \xi = \sum_{i=1}^{p} \lambda_{m,i} \xi_i , \quad m = 1, 2, \ldots, M ,$$

where $\lambda_m = (\lambda_{m,1}, \lambda_{m,2}, \ldots, \lambda_{m,p})^T$ are randomly chosen.

If $M$ is sufficiently large (at least $M > p$), we obtain approximately that if $\eta_1, \eta_2, \ldots, \eta_M$ are all univariate normal, then the random vector $\xi = (\xi_1, \xi_2, \ldots, \xi_p)^T$ is distributed according to multivariate normal distribution. Thus the null hypothesis converts almost equivalently to $H_0: \eta_1, \eta_2, \ldots, \eta_M$ are all univariate normal. For each $\eta_m$ ($m = 1, 2, \ldots, M$), with the result presented in the previous section, we implement hypothesis tests for univariate normality, which we call subtests, and decide whether $\eta_m$ is distributed according to univariate normal distribution. Note that we can not conclude that $\xi$ is not normal even if there exit subtests which decide $\eta_m$ is not normal, since the decision of a subtest is not correct with probability 1. It only guarantees that the probability of type I error in the subtest is controlled by the subtest’s significance level $\alpha_m$.

Taking into consideration of (1), for the $m$th subtest we have

$$P_i(e) = P[\text{Decide } \eta_m \text{ is not normal} \mid \eta_m \text{ is normal}] = \alpha_m.$$  (4)

and

$$P[\text{Decide } \eta_m \text{ is normal} \mid \eta_m \text{ is normal}]$$

$$= 1 - P_i(e) = 1 - \alpha_m .$$  (5)

For convenience, the significance levels of all subtests are chosen to be the same, and denoted by $\alpha$.

In the main hypothesis test that $\eta_1, \eta_2, \ldots, \eta_M$ are all univariate normal, there are always some subtests which decide the corresponding $\eta_m$ is not normal. We thus choose the number of the subtests which decide their $\eta_m$ is not normal as the test statistic, and denote it as $K$ ($0 \leq K \leq M$) in the rest of this paper. Given the significance level $\alpha$ of the main test, we need to determine the critical value $\hat{k}$. If $K \geq \hat{k}$, we reject the hypothesis that $\eta_1, \eta_2, \ldots, \eta_M$ are all normal, and consequently decide that $\xi$ is not distributed according to multivariate normal distribution. We also know that the probability of type I error of the test is controlled by $\alpha$,

$$P_i(e) = P[\text{Decide } \xi \text{ is not normal} \mid \xi \text{ is normal}]$$

$$= P[\text{Decide not all } \eta_1, \eta_2, \ldots, \eta_M \text{ are normal} \mid \eta_1, \eta_2, \ldots, \eta_M \text{ are normal}]$$

$$= \alpha .$$  (6)

In order to compute the critical value $\hat{k}$, we use the following equalities.

$$\alpha = P[\text{Decide not all } \eta_1, \eta_2, \ldots, \eta_M \text{ are normal} \mid \eta_1, \eta_2, \ldots, \eta_M \text{ are normal}]$$

$$= P[K \geq \hat{k} \mid \eta_1, \eta_2, \ldots, \eta_M \text{ are all normal}]$$

$$= \sum_{i=0}^{\hat{k}} P[K = i \mid \eta_1, \eta_2, \ldots, \eta_M \text{ are all normal}] .$$  (7)

Note that

$$P[K = \hat{k} + i \mid \eta_1, \eta_2, \ldots, \eta_M \text{ are all normal}]$$

$$= C_{\hat{k}}^{i} \alpha^{i} (1 - \alpha)^{M-(i+1)} .$$  (8)

Where $\alpha_0$ is the significance level of the subtests. Combining (7) and (8), we obtain,

$$\sum_{i=0}^{\hat{k}} C_{\hat{k}}^{i} \alpha^{i} (1 - \alpha)^{M-(i+1)} = \alpha .$$  (9)

For convenience, let the critical value $\hat{k}$ be the minimal integer satisfying

$$\sum_{i=0}^{\hat{k}} C_{\hat{k}}^{i} \alpha^{i} (1 - \alpha)^{M-(i+1)} \leq \alpha .$$  (10)

In fact, since $\alpha$ is a small number (0.01 or 0.05), it is sufficient to compute the first a few terms in the summation.

So far, we have completely constructed a hypothesis test, which can decide whether a random vector is distributed according to multivariate normal distribution.

### 4. RESULTS

We applied the hypothesis test to decide whether the prior distributions of the mean vectors of the MFCC are multivariate normal. The speech database is provided by national 863 High-
Tech project of P.R.C. The number of the linear combinations \( M \) is chosen to be 500, when significance level is set as 0.01 — which means that the probability of a false rejection is no more than 1% — the critical value \( K \) is 12. Experimental results show that 611 out of 830 valid states reject the normality assumption. The histogram of the number of subtests, which decide their linear combinations are non-normal, is given in Fig1.

Fig 1  Histogram of the number of subtests that decide their linear combination is not normal.

5. CONCLUSION

This paper presents an algorithm for testing whether a random vector is distributed to multivariate normal distribution. When applied to the Bayesian approach of speaker adaptation, experimental results show that the prior distributions of the mean of MFCC are far from normality. This implies that better choice of the prior form may improve the adaptation result.

6. REFERENCES