Reprinted from

Seventh International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Range, Forest and Wetlands Assessment

June 23 - 26, 1981

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

THE EXPANSION OF THE PROBABILITY DENSITY FUNCTION TO NON-GAUSSIAN DISTRIBUTION

MINORU AKIYAMA

Geographical Survey Institute
Ibaraki-ken, Japan

I. ABSTRACT

In the most of all statistical approaches in pattern recognition theories in remote sensing it is assumed that each probability density function of pattern classes can be approximated by the Gaussian probability density function. However, this assumption is not always appropriate in practice. The exact shape of class probability density function is supposed to be expressed as an original histogram. And if the shape of the histogram is largely different from the Gaussian function the classification results might include large error.

Therefore, there seems no need to persist in Gaussian probability density function as the only representation of class histograms. In other words, if there are other functions which can approximate the original histograms more accurately than the Gaussian function can, we would better to adopt one of those functions as a representation of a pattern class histogram.

From this point of view, a probability density function was expanded by adding another parameter to the Gaussian function so that it can approximate histograms more flexibly and still can include the Gaussian function itself as a special case.

The expanded function used here is a non-symmetric Gaussian function which has two independent standard deviations for each side of the mode so that it can approximate the anti-symmetricity of class histogram.

In this paper some characteristics of the non-symmetric Gaussian probability density function were studied. Then the fitness to the original histogram was examined by chi-square test and compared with that of the conventional symmetric Gaussian function. The comparison between symmetric and non-symmetric function was accomplished also on the results of a test run.

II. NON-SYMMETRIC GAUSSIAN PROBABILITY DENSITY FUNCTION

In the discriminate procedure of maximum likelihood classification, the most important point as a good approximate function is how accurate it can define the decision boundary rather than how accurate it can trace the fine feature of the original histogram.

The position of the decision boundary should be defined as the intersection of the histograms of two classes. Even though we limit our considerations to unimodal classes, actual histogram can be characterized in various ways. However, it seems natural to introduce anti-symmetricity as the third parameter besides mean and standard deviation. Because the position of the decision boundary depends on the positions of the modes and the decreasing ratios of each class, and the decreasing ratio of the side where intersection may exist is not necessarily the same as that of the other side.

From the above mentioned point of view, non-symmetric Gaussian probability density function was considered. This function has two independent standard deviations for each side of the mode and will be equal to the Gaussian function only when the two standard deviations are equal to each other.

In one dimensional case, the function is defined as

$$P(x) = \frac{2}{\sqrt{2\pi}(\sigma_l + \sigma_r)} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad \begin{cases} \sigma = \sigma_l \text{ for } x \leq \mu \\ \sigma = \sigma_r \text{ for } x > \mu \end{cases} \tag{1}$$

where μ is a mode rather than a mean, σ_l and σ_r are the left and the right standard deviations respectively.

In order to obtain the set of three parameters μ, σ_l and σ_r from original data, let us discuss some characteristics of the non-symmetric Gaussian probability density function.

At first, from the definition of the function (1) we easily obtain some integration formulae as

$$\int_{-\infty}^{\infty} P(x)dx = 1 \tag{2}$$

$$\int_{-\infty}^{\infty} xP(x)dx = \mu - \sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) \tag{3}$$

$$\int_{-\infty}^{\infty} x^2 P(x)dx = \mu^2 - 2\mu\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) + (\sigma_l^2 - \sigma_l\sigma_r + \sigma_r^2) \tag{4}$$

$$\int_{-\infty}^{\infty} x^3 P(x)dx = \mu^3 - 3\mu^2\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) + 3\mu(\sigma_l^2 - \sigma_l\sigma_r + \sigma_r^2) - 2\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r)(\sigma_l^2 + \sigma_r^2) \tag{5}$$

These integrations are corresponding to sum, square sum and cubic sum of the original data. Let Σ_1, Σ_2 and Σ_3 be average, square average and cubic average respectively. Then we obtain

$$\Sigma_1 = \frac{1}{q}\sum_{i=1}^{q} x_i = \int_{-\infty}^{\infty} xP(x)dx \tag{6}$$

$$\Sigma_2 = \frac{1}{q}\sum_{i=1}^{q} x_i^2 = \int_{-\infty}^{\infty} x^2P(x)dx \tag{7}$$

$$\Sigma_3 = \frac{1}{q}\sum_{i=1}^{q} x_i^3 = \int_{-\infty}^{\infty} x^3P(x)dx \tag{8}$$

Let $\delta = \mu - \Sigma_1$, then we obtain the equation (see Appendix A)

$$(\pi - 3)\delta^3 - (\Sigma_2 - \Sigma_1^2) - (\Sigma_3 - 3\Sigma_1\Sigma_2 + 2\Sigma_1^3) = 0 \tag{9}$$

and

$$\mu = \Sigma_1 + \delta \tag{10}$$

$$\sigma_l = \left\{\Sigma_2 - \Sigma_1^2 + (1 - \frac{3}{8}\pi)\delta^2\right\}^{1/2} + \sqrt{\frac{\pi}{8}}\,\delta \tag{11}$$

$$\sigma_r = \left\{\Sigma_2 - \Sigma_1^2 + (1 - \frac{3}{8}\pi)\delta^2\right\}^{1/2} - \sqrt{\frac{\pi}{8}}\,\delta \tag{12}$$

Thus, the parameters μ, σ_l and σ_r can be obtained from Σ_1, Σ_2, Σ_3 and δ. And δ is a solution of the equation (9) which is the third order equation. Here let us discuss a few points about this equation. The coefficient of the third order term $(\pi - 3)$ is obviously positive and the coefficient of the first order term $-(\Sigma_2 - \Sigma_1^2)$ is always negative because the value $(\Sigma_2 - \Sigma_1^2)$ corresponds to the conventional variance σ^2. Therefore, the function of the left-hand side of the equation always has maximum and minimum value. And through some considerations on the meaning of δ which is the difference between mean and mode, the solution we need should be exist between the maximum and the minimum points.

However, the equation (9) has such a solution only when the maximum value is positive and the minimum value is negative. In fact, there are some cases which the shape of the histogram is extremely different from the Gaussian type that we can not obtain the correct solution for δ and can not define the non-symmetric Gaussian function.

In order to avoid this difficulty we took a different definition for δ when the equation (9) does not give an adequate solution. More details will be mentioned in Appendix B.

Fig. 1 shows an example of the non-symmetric Gaussian function together with the corresponding original histogram and regular symmetric Gaussian function.

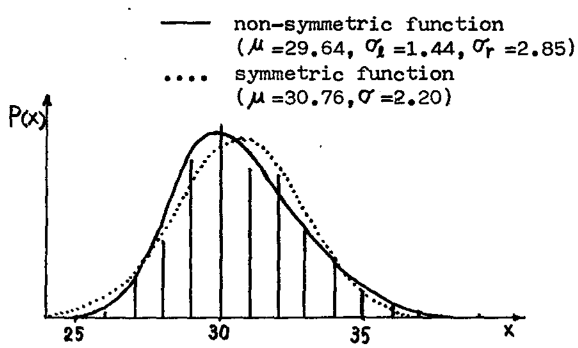

Fig. 1 Symmetric and Non-symmetric Gaussian probability density function

In this case the fitness of the non-symmetric function to the original histogram is more than 80% according to chi-square test. On instead, the fitness of the symmetric function is less than 30%. This is only an example, however, most of other classes also shows the significant improvement in fitness.

Fig. 2 of the next page shows an example of the comparison of decision boundaries derived from non-symmetric and symmetric functions. This example shows only around the decision boundaries, however, it clearly shows that the non-symmetric function has a effect to move the decision boundary close to where is expected from the original histograms.

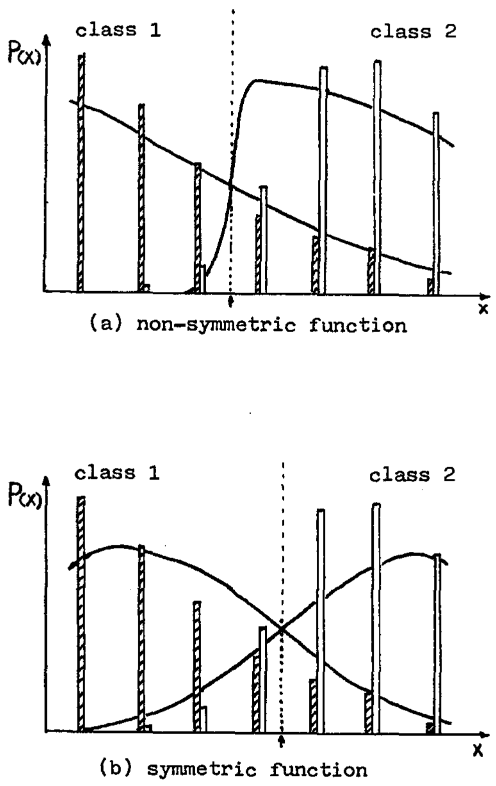

Fig. 2 Decision boundary from non-symmetric and symmetric function

III. EXPANSION TO MULTI DIMENSION

We have discussed non-symmetric Gaussian function for only one-dimensional case so far. Now we must expand this function to multi dimension.

Before considering non-symmetric function, let us refer to the symmetric function. The multi dimensional Gaussian probability density function is given by

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{u})^T V^{-1}(\vec{x}-\vec{u})\right\} \qquad (13)$$

where $\vec{x}$ is a data vector, $\vec{u}$ is the mean vector, V is the covariance matrix, $|V|$ is the determinant of V, V^{-1} is the inverse matrix of V and $(\vec{x}-\vec{u})^T$ is the transpose of vector $(\vec{x}-\vec{u})$. Let Γ be the correlation matrix and S be the diagonal matrix with components of the set of standard deviations, that is

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{pmatrix} \qquad \left(\text{where } \gamma_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\cdot\sigma_{jj}}}\right) \qquad (14)$$

$$S = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix} \qquad (15)$$

then we obtain

$$V = S\Gamma S \qquad (16)$$

$$|V|^{1/2} = (|S||\Gamma||S|)^{1/2} = |\Gamma|^{1/2} \prod_{i=1}^{n} \sigma_i \qquad (17)$$

$$V^{-1} = S^{-1}\Gamma^{-1}S^{-1} \qquad (18)$$

therefore, (13) will be rewritten as

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2}|\Gamma|^{1/2}\prod_{i=1}^{n}\sigma_i} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{u})^T S^{-1}\Gamma^{-1}S^{-1}(\vec{x}-\vec{u})\right\} \qquad (19)$$

As analogy to one dimensional case, the multi dimensional non-symmetric Gaussian function can be given as

$$P(\vec{x}) = \frac{2^n}{(2\pi)^{n/2}|\Gamma|^{1/2}\prod_{i=1}^{n}(\sigma_{l_i}+\sigma_{r_i})} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{u})^T S^{-1}\Gamma^{-1}S^{-1}(\vec{x}-\vec{u})\right\} \qquad (20)$$

where

$$S = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix} \qquad \begin{cases} \sigma_i = \sigma_{l_i} & \text{for } x_i \le \mu_i \\ \sigma_i = \sigma_{r_i} & \text{for } x_i > \mu_i \end{cases} \qquad (21)$$

and $\vec{u}$ is the mode vector.

In this replacement, we assumed that the correlation matrix Γ is constant in whole space regardless of the change from σ_l to σ_r. This assumption is not a priori, however, it is not unreasonable either. In fact, it is almost impossible to define correlation matrices for each subspaces according to each combination of left or right standard deviations for each channels, because of the restricted number of data and large number of memory location requirement.

IV. COMPARISON WITH THE CONVENTIONAL METHOD

A. COMPUTATION TIME

For this study some LARSYS programs were modified temporarily so that they can deal with the non-symmetric Gaussian function. The modified functions are "CLUSTER" function which is for clustering and calculation of statistics for each cluster, "STATISTICS" function which calculate statistics for classes corresponding to indicated region of data, "MERGESTATISTICS" function which merge statistics derived from "CLUSTER" and/or "STATISTICS" function and "CLASSIFYPOINT" function which is point classification by means of maximum likelihood method.

As far as computation time is concerned, "STATISTICS" and MERGESTATISTICS" functions can be negligible inthe whole procedure of classification.

"CLUSTER" function needs longer time, however, the increase of computation time of this function caused by modification is negligible compared with the total time of "CLUSTER" function. The problem lies in "CLASSIFYPOINT" function. The increase of computation time is up to 50% of the total time of "CLASSIFYPOINT" function. However, this increase of computation time can be reduced because the subroutine for the main process of "CLASSIFYPOINT" function in the modified version was written in FORTRAN as compared with the original version written in ASSEMBLER.

B. TEST RUN

In order to see the effect of this approach, a test run was accomplished. The data used in this study is a part of LANDSAT MSS data of scene 1321-15595 collected on June 9, 1973 at 9:59 am. The object area includes Monroe Reservoir and Bloomington, Indiana. We selected these data for a test run simply because these data has been studied for many years as a standard data set for education and program tests in LARS. Futhermore, some other informations including ground truth data are also available together with the data.

The ground truth data are such that more than three hundred test areas of 3X3 points data are assigned to five categories as urban, agricaltural fields, forest, water and cloud. Therefore, after classified the data into many classes, those classes were grouped into those five categories.

Classification study was accomplished in combination of supervised and unsupervised methods. Moreover, the comparison of symmetric and non-symmetric functions were accomplished on the results of classifications in some different numbers of classes.

Through this study, the following results were obtained.

At first, this method can improve the trainning field performance in supervised approach, however, it does not give a remarkable effect in unsupervised approach. The cause of this fact is supposed to be in that each cluster does not necessarily correspond to the actual pattern class.

Secondary, it gives more effect when the number of pattern classes is small. When the number of pattern classes is large, the standard deviation of each class should be small. Therefore, they makes little difference in the position of decision boundary even though the shape of histogram of a class is very anti-symmetric.

Finally, non-symmetric function often cut the harf side of the feature space very sharply. Therefore, it may happen that those data which include more than two classes can be assigned to a very different class.

V. APPENDIX

A. INTRODUCTION OF EQUATION (9)

The following equations are obtained on reffering to equations (3) through (8).

$$\Sigma_1 = \mu - \sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) \tag{22}$$

$$\Sigma_2 = \mu^2 - 2\mu\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) + (\sigma_l^2 - \sigma_l\sigma_r + \sigma_r^2) \tag{23}$$

$$\Sigma_3 = \mu^3 - 3\mu^2\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) + 3\mu(\sigma_l^2 - \sigma_l\sigma_r + \sigma_r^2) - 2\sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r)(\sigma_l^2 + \sigma_r^2) \tag{24}$$

Here, let δ and σ be

$$\delta = \sqrt{\frac{2}{\pi}}(\sigma_l - \sigma_r) \tag{25}$$

$$\sigma = \frac{1}{2}(\sigma_l + \sigma_r) \tag{26}$$

then equation (22)-(24) are rewritten as

$$\Sigma_1 = \mu - \delta \tag{27}$$

$$\Sigma_2 = \mu^2 - 2\mu\delta + \frac{3\pi}{8}\delta^2 + \sigma^2 \tag{28}$$

$$\Sigma_3 = \mu^3 - 3\mu^2\delta + 3\mu\left(\frac{3\pi}{8}\delta^2 + \sigma^2\right) - 2\delta\left(\frac{\pi}{4}\delta^2 + 2\sigma^2\right) \tag{29}$$

From (27) and (28) we obtain

$$\Sigma_2 - \Sigma_1^2 = \sigma^2 + \left(\frac{3\pi}{8} - 1\right)\delta^2 \tag{30}$$

or

$$\sigma^2 = \Sigma_2 - \Sigma_1^2 - \left(\frac{3\pi}{8} - 1\right)\delta^2 \tag{31}$$

and from (27)

$$\mu = \Sigma_1 + \delta \tag{32}$$

By substituting (31) and (32) into (29) then we obtain

$$\Sigma_3 = \Sigma_1^3 + (\pi - 3)\delta^3 + 3\Sigma_1(\Sigma_2 - \Sigma_1^2) - (\Sigma_2 - \Sigma_1^2)\delta \tag{33}$$

(33) is rearranged as

$$(\pi - 3)\delta^3 - (\Sigma_2 - \Sigma_1^2)\delta - (\Sigma_3 - 3\Sigma_1\Sigma_2 + 2\Sigma_1^3) = 0 \tag{34}$$

By reffering (25) and (26) we obtain

$$\sigma_l = \sigma + \sqrt{\frac{\pi}{8}}\delta \tag{35}$$

$$\sigma_r = \sigma - \sqrt{\frac{\pi}{8}}\delta \tag{36}$$

By substituting (31) we obtain

$$\sigma_\ell = \left\{\Sigma_2 - \Sigma_1^2 + (1 - \frac{3}{8}\pi)\delta^2\right\}^{1/2} + \sqrt{\frac{\pi}{8}}\,\delta \quad (37)$$

$$\sigma_r = \left\{\Sigma_2 - \Sigma_1^2 + (1 - \frac{3}{8}\pi)\delta^2\right\}^{1/2} - \sqrt{\frac{\pi}{8}}\,\delta \quad (38)$$

and

$$\mu = \Sigma_1 + \delta \quad (32)$$

B. SOLUTION OF EQUATION (9)

Standard deviations σ_ℓ and σ_r should be positive, however, it never happens that both of them will be negative at a time, because of (37) and (38). Therefore, the product of σ_ℓ and σ_r should be positive. That is

$$\sigma_\ell \sigma_r = \hat{\sigma}^2 - (\frac{\pi}{2} - 1)\delta^2 > 0 \quad (39)$$

$$\therefore \; \delta^2 < \frac{2}{\pi - 2}\hat{\sigma}^2 \quad (40)$$

where

$$\hat{\sigma}^2 = \Sigma_2 - \Sigma_1^2 \quad (41)$$

Moreover, we obtain

$$\hat{\sigma}^2 - (\frac{3}{8}\pi - 1)\delta^2 > \hat{\sigma}^2 - (\frac{3}{8}\pi - 1)\frac{2}{\pi - 2}\hat{\sigma}^2$$
$$= \frac{\pi}{4(\pi - 2)}\hat{\sigma}^2 > 0 \quad (42)$$

Therefore, if the inequality (40) is satisfied it is also guaranteed that both σ_ℓ and σ_r are real numbers.

Let $f(\delta)$ be the left hand side of the equation (34). That is

$$f(\delta) = (\pi - 3)\delta^3 - \hat{\sigma}^2\delta - M_3 \quad (43)$$

where

$$M_3 = \Sigma_3 - 3\Sigma_1\Sigma_2 + 2\Sigma_1^3 \quad (44)$$

then

$$f'(\delta) = 3(\pi - 3)\delta^2 - \hat{\sigma}^2 \quad (45)$$

Let $\pm\Delta$ ($\Delta > 0$) be the maximum and minimum points.

$$\pm\Delta = \pm\frac{\hat{\sigma}}{\sqrt{3(\pi - 3)}} \quad (46)$$

and then,

$$\frac{\hat{\sigma}^2}{3(\pi - 3)} > \frac{2}{\pi - 2}\hat{\sigma}^2 \quad (47)$$

Therefore, any δ which satisfy (40) will be exist between maximum and minimum points. Then in the region

$$-\sqrt{\frac{2}{\pi - 2}}\,\hat{\sigma} < \delta < \sqrt{\frac{2}{\pi - 2}}\,\hat{\sigma} \quad (48)$$

$f(\delta)$ is a simply increasing function. Therefore, the condition that the equation (34) has a solution in the region can be expressed as

$$f(-\sqrt{\frac{2}{\pi - 2}}\,\hat{\sigma})\; f(\sqrt{\frac{2}{\pi - 2}}\,\hat{\sigma}) < 0 \quad (49)$$

$$M_3^2 < \frac{2(4 - \pi)^2}{(\pi - 2)^3}\hat{\sigma}^6 \quad (50)$$

Inequality (50) gives the condition to satisfy (40). Therefore, when the condition (50) is not satisfied we should define δ in another way. Here, we took the following definition.

$$\delta = 1.2\,\hat{\sigma} \quad \text{for } M_3 < 0$$
$$\delta = -1.2\,\hat{\sigma} \quad \text{for } M_3 > 0 \quad (51)$$

The value 1.2 was chosen on reffering to the coefficient of the right hand member of inequality (40). That is

$$\sqrt{\frac{2}{\pi - 2}} = 1.3236\ldots \quad (52)$$

Then the inequality (50) was also changed as

$$M_3^2 < 0.913\,\hat{\sigma}^6 \quad (53)$$

so that it can match with the definition (51).

In other words, when (53) is satisfied δ is obtained by solving the equation (34), if not, δ is obtained from (51).

AUTHOR BIOGRAPHICAL DATA

Minoru Akiyama. Technical official of Geographical Survey Institute, Ministry of Construction in Japan. He has worked on remote sensing and photogrametry since 1974. Currently he is studying data processing and analysis at Purdue/LARS as a visiting scientist.