Теория сигналов и систем

УДК 519.213

К.П. Пилипенко

Умови приналежності двокомпонентної суміші розподілів до одного типу

Рассмотрен общий случай двухкомпонентной гауссовской смеси распределений. Показано, что варьируя набором параметров составляющих смеси можно получить большое семейство распределений. Получено условие принадлежности функций распределения смеси одному типу. Определены требования к параметрам составляющих смеси, при выполнении которых смеси являются распределениями одного типа. Выведены формулы перехода от случайной величины с произвольными значениями математического ожидания и дисперсии к стандартной случайной величине с нулевым математическим ожиданием и единичной дисперсией.

A general case the two-component Gaussian mixture is examined. It is shown that by varying the parameters of components of the mixture a large family of distributions can be obtained. A criterion for the distribution function is found, defining the mixtures’ belonging to one type. The conditions, which the component parameters of uniform mixtures should meet, are determined. Formulas for the transition from a random variable with arbitrary values of the expectation and variance to the standard random variable with zero mean and unit variance were derived.

Ключевые слова: *случайная величина, гауссовская смесь, плотность вероятностей.*

Вступ

В області статистичної обробки процесів найбільш поширеною моделлю для опису випадкових сигналів і завад є гаусівський розподіл. Проте розподіл багатьох процесів істотно відрізняється від гаусівського. До таких процесів відносяться мовні та акустичні флуктуаційні сигнали [1, 2]. Ряд застосувань сумішей розподілів в області нелінійної фільтрації сигналів і аналізу негаусівських шумових процесів наведено в [3], а в роботі [1] в якості статистичної моделі мовного сигналу запропоновано використовувати частковий випадок суміші розподілів. Таким чином, актуальною є задача дослідження негаусівських процесів, до числа яких відносяться суміші розподілів.

Дискретна суміш розподілів визначається наступним чином [4, 5]:

$$F(x) = \sum_k d_k F_k(x), \tag{1}$$

де $F_k(x)$ – деякі функції розподілу, які, як правило, є однаковими; d_k – вагові коефіцієнти суміші, що задовольняють умові:

$$\begin{cases} d_k \ge 0, \\ \sum\limits_k d_k = 1. \end{cases} \tag{2}$$

Деякі задачі, що стосуються окремих випадків сумішей розподілів розглянуті в роботах [6, 7].

Мета роботи – визначення умов приналежності до одного типу законів розподілу двокомпонентних гаусівских сумішей.

Приналежність функції розподілу суміші до одного типу

Розглянемо випадкову величину ξ з функцією розподілу $F(x)$, що являє собою двокомпонентну суміш (1)

$$F(x) = d_1F_1(x) + d_2F_2(x), \tag{3}$$

складові $F_1(x)$ та $F_2(x)$ якої є гаусівськими функціями розподілу

$$F_k(x) = 0{,}5 + \Phi_0\left(\frac{x-m_k}{\sigma_k}\right),\quad k = 1{,}2,$$

де $\Phi_0(x)$ – функція Лапласа,

$$\Phi_0(x) = \int_0^x \varphi(x)dx,$$

а коефіцієнти d_1 і d_2 задовольняють умовам (2)

Щільність ймовірностей $p(x)$ суміші (3) описується виразом

$$p(x) = \frac{d_1}{\sigma_1}\varphi\left(\frac{x-m_1}{\sigma_1}\right) + \frac{d_2}{\sigma_2}\varphi\left(\frac{x-m_2}{\sigma_2}\right), \tag{4}$$

де $\varphi(x)$ – щільність ймовірностей стандартної гаусівської випадкової величини,

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

З формули (4) легко отримати математичне сподівання та дисперсію суміші розподілів:

$$m=d_1m_1+d_2m_2, \quad (5)$$

$$\sigma^2=d_1d_2\left(m_1-m_2\right)^2+d_1\sigma_1^2+d_2\sigma_2^2, \quad (6)$$

де m_1, m_2, σ_1^2 и σ_2^2 – математичні сподівання та дисперсії складових суміші.

У зв'язку з тим, що, варіюючи набором параметрів $\{d_1,d_2,m_1,m_2,\sigma_1,\sigma_2\}$, які визначають суміш, можна отримати велике сімейство розподілів, що визначаються формулами (3) і (4) (рис. 1, 2), виникає задача знаходження умов, за яких розподіли (3) належать до одного і того ж типу.

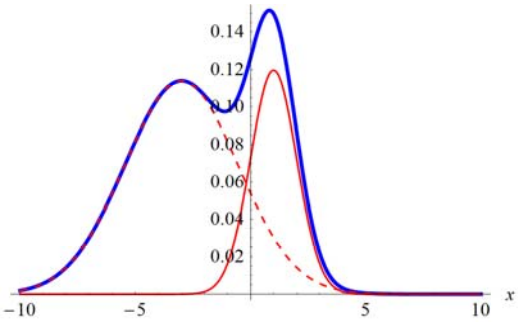

Рис. 1. Суміш розподілів з параметрами, $d_1=0{,}7$, $d_2=0{,}3$, $\sigma_1=2{,}45$, $\sigma_2=1$, $m_1=-3$, $m_2=1$, **—** – $p(x)$, - - – $d_1p_1(x)$, — – $d_2p_2(x)$.

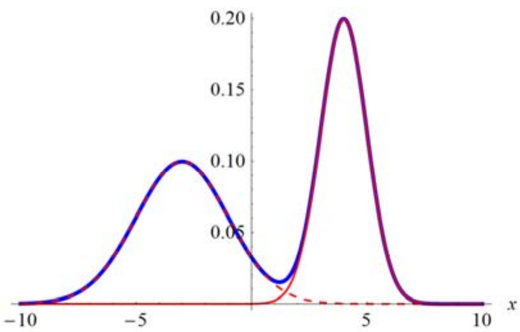

Рис. 2. Суміш розподілів з параметрами $d_1=0{,}5$, $d_2=0{,}5$, $\sigma_1=2$, $\sigma_2=1$, $m_1=-3$, $m_2=4$, **—** – $p(x)$, - - – $d_1p_1(x)$, — – $d_2p_2(x)$.

Згідно [4], для того щоб дві випадкові величини ξ_1 і ξ_2 з функціями розподілу належали до одного і того ж типу, необхідно виконання умови

$$F_2(x)=F_1\left(\frac{x-b}{a}\right), \quad (7)$$

де $a>0$, $b\in(-\infty,\infty)$.

З виразу (7) робимо висновок, що випадкові величини ξ_1 і ξ_2 належать до одного і того ж типу, якщо вони пов'язані лінійною залежністю, а саме

$$\xi_2=a\xi_1+b.$$

Знайдемо умови, за яких випадкові величини ξ_1 і ξ_2 функції розподілу яких визначаються виразом (3), належать до одного типу, але мають різні параметри.

Функції розподілу випадкових величин ξ_1 і ξ_2, відповідно, $F_1(x)$ і $F_2(x)$, запишуться наступним чином (3)

$$F_1(x)=0{,}5+d_{11}\Phi_0\left(\frac{x-m_{11}}{\sigma_{11}}\right)+d_{12}\Phi_0\left(\frac{x-m_{12}}{\sigma_{12}}\right), \quad (8)$$

$$F_2(x)=0{,}5+d_{21}\Phi_0\left(\frac{x-m_{21}}{\sigma_{21}}\right)+d_{22}\Phi_0\left(\frac{x-m_{22}}{\sigma_{22}}\right). \quad (9)$$

Очевидно, що математичні сподівання та дисперсії випадкових величин ξ_1 і ξ_2 дорівнюють , відповідно (5), (6):

$$m_k=M\{\xi_k\}=d_{k1}m_{k1}+d_{k2}m_{k2}, \quad (10)$$

$$\sigma_k^2=d_{k1}d_{k2}\left(m_{k1}-m_{k2}\right)^2+d_{k1}\sigma_{k1}^2+d_{k2}\sigma_{k2}^2, \quad (11)$$

де $k=1{,}2$.

Відомо, що дві випадкові величини ξ_1 і ξ_2 належать до одного типу за виконання умов:

$$\frac{\xi_1-m_1}{\sigma_1}=\frac{\xi_2-m_2}{\sigma_2}=\xi_0. \quad (12)$$

З виразу (12) отримуємо

$$\xi_2=\xi_0\sigma_2+m_2=\frac{\xi_1-m_1}{\sigma_1}\sigma_2+m_2=$$

$$=\frac{\sigma_2}{\sigma_1}\xi_1-\frac{\sigma_2}{\sigma_1}m_1+m_2.$$

З останнього виразу видно, що випадкові величини ξ_2 і ξ_1 пов'язані лінійною залежністю

$$y=G(x)=\frac{\sigma_2}{\sigma_1}x-\frac{\sigma_2}{\sigma_1}m_1+m_2, \quad (13)$$

яка є монотонно зростаючою функцією.

Відомо [8], що якщо випадкова величина ξ_2 отримана з випадкової величини ξ_1 за допомогою монотонно зростаючої функції $y=G(x)$, то функція розподілу $F_2(y)$ виражається через функцію розподілу $F_1(x)$ наступним чином:

$$F_1(y) = F_1\left[G^{-1}(y)\right],$$

де $x = G^{-1}(y)$ – функція, обернена до $G(x)$.

В даному випадку із (13) отримуємо обернену функцію

$$x = \frac{\sigma_1}{\sigma_2}(y - m_2) + m_1 \quad (14)$$

Підставляючи (14) в (8), отримуємо остаточну формулу для функції розподілу $F_2(y)$ випадкової величини ξ_2, яка належить до того ж типу, що і випадкова величина ξ_1:

$$F_2(y) = F_1\left[\frac{\sigma_1}{\sigma_2}(y - m_2) + m_1\right] =$$

$$= 0{,}5 + d_{11}\Phi_0\left(\frac{\frac{\sigma_1}{\sigma_2}(y - m_2) + m_1 - m_{11}}{\sigma_{11}}\right) +$$

$$+ d_{12}\Phi_0\left(\frac{\frac{\sigma_1}{\sigma_2}(y - m_2) + m_1 - m_{12}}{\sigma_{12}}\right). \quad (15)$$

Вимоги до параметрів сумішей одного типу

Знайдемо загальні умови, яким повинні задовольняти параметри розподілів (8) і (9), щоб відповідні їм випадкові величини ξ_1 і ξ_2 належали до одного типу. Для цього приведемо функцію розподілу (15) до форми (9). Видно, що для збігу виразів (15) і (9) необхідно виконання умов

$$\begin{cases} d_{11} = d_{21}; \\ d_{12} = d_{22}, \end{cases} \quad (16)$$

$$\begin{cases} \dfrac{\sigma_1}{\sigma_2\sigma_{11}} = \dfrac{1}{\sigma_{21}}; \\ \dfrac{\sigma_1}{\sigma_2\sigma_{12}} = \dfrac{1}{\sigma_{22}}, \end{cases} \quad (17)$$

$$\begin{cases} \dfrac{-m_1 + m_{11} + m_2\frac{\sigma_1}{\sigma_2}}{\sigma_{11}} = \dfrac{m_{21}}{\sigma_{21}}; \\ \dfrac{-m_1 + m_{12} + m_2\frac{\sigma_1}{\sigma_2}}{\sigma_{12}} = \dfrac{m_{22}}{\sigma_{22}}. \end{cases} \quad (18)$$

Перепишемо формули (17) наступним чином

$$\begin{cases} \dfrac{\sigma_1}{\sigma_2} = \dfrac{\sigma_{11}}{\sigma_{21}}; \\ \dfrac{\sigma_1}{\sigma_2} = \dfrac{\sigma_{12}}{\sigma_{22}}, \end{cases} \quad (19)$$

і, прирівнявши праві частини виразів (19), одержимо умову, якій повинні задовольняти дисперсії складових компонент розподілів одного типу:

$$\frac{\sigma_{11}}{\sigma_{21}} = \frac{\sigma_{12}}{\sigma_{22}}, \quad (20)$$

або

$$\frac{\sigma_{11}}{\sigma_{12}} = \frac{\sigma_{21}}{\sigma_{22}}. \quad (21)$$

Перезапишемо тепер вирази (18):

$$\begin{cases} m_2\dfrac{\sigma_1}{\sigma_2} - m_1 = \dfrac{\sigma_{11}}{\sigma_{21}}m_{21} - m_{11}; \\ m_2\dfrac{\sigma_1}{\sigma_2} - m_1 = \dfrac{\sigma_{12}}{\sigma_{22}}m_{22} - m_{12}, \end{cases}$$

прирівнявши праві частини, і, враховуючи, (20) і (21) отримуємо умови для математичних сподівань складових суміші:

$$\frac{\sigma_{11}}{\sigma_{21}}(m_{21} - m_{22}) = m_{11} - m_{12} \quad (22)$$

або, якщо $m_{21} \neq m_{22}$, маємо

$$\frac{\sigma_{11}}{\sigma_{21}} = \frac{\sigma_{12}}{\sigma_{22}} = \frac{m_{11} - m_{12}}{m_{21} - m_{22}}. \quad (23)$$

Із (22) випливає, що якщо $p_1(x)$ є одновершинною симетричною сумішшю розподілів $(m_{11} = m_{12})$, то і щільність ймовірностей того ж типу $p_2(x)$ також є одновершинною симетричною сумішшю $(m_{21} = m_{22})$.

Таким чином, для того, щоб два закони розподілу (8) і (9) належали до одного типу, параметри складових розподілів повинні задовольняти умовам (16) і (23), а, в разі рівності математичних очікувань, належність до одного типу визначається виразами (16) і (21).

Як приклад розглянемо дві щільності ймовірностей $p_1(x)$ та $p_2(x)$, які описуються виразом (4). Параметри щільності ймовірностей $p_1(x)$:

$$d_{11} = 0{,}3;\ d_{12} = 0{,}7;\ \sigma_{11} = 1;\ \sigma_{12} = 3;$$
$$m_{11} = 0;\ m_{12} = 2, \quad (24)$$

параметри щільності ймовірностей $p_2(x)$:

$d_{21} = 0{,}3$; $d_{22} = 0{,}7$; $\sigma_{21} = 0{,}5$; $\sigma_{22} = 1{,}5$;

$m_{21} = 1$; $m_{22} = 2$. (25)

Легко помітити, що параметри (24) і (25) задовольняють умовам (16) і (23), таким чином можна зробити висновок, що щільності ймовірностей $p_1(x)$ і $p_2(x)$ належать до одного типу. Обчислимо математичні очікування й дисперсії, відповідно до (10) та (11):

$$m_1 = 1{,}4;\ \sigma_1 = 2{,}73;$$

$$m_2 = 1{,}7;\ \sigma_2 = 1{,}36.$$

Графіки відповідних щільностей ймовірностей зображені на рис. 3, 4.

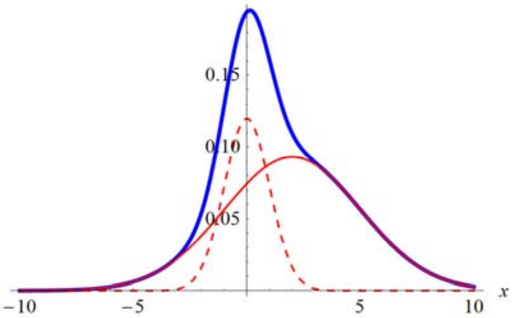

Рис. 3. Суміш розподілів з параметрами $d_{11} = 0{,}3$, $d_{12} = 0{,}7$, $\sigma_{11} = 1$, $\sigma_{12} = 3$, $m_{11} = 0$, $m_{12} = 2$, — $p_1(x)$, - - $d_{11}p_{11}(x)$, — $d_{12}p_{12}(x)$.

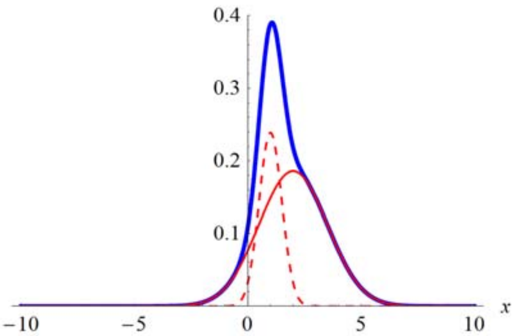

Рис. 4. Суміш розподілів з параметрами $d_{21} = 0{,}3$, $d_{22} = 0{,}7$, $\sigma_{21} = 0{,}5$, $\sigma_{22} = 1{,}5$, $m_{21} = 1$, $m_{22} = 2$, — $p_2(x)$, - - $d_{21}p_{21}(x)$, — $d_{22}p_{22}(x)$.

Формули переходу

Приведемо формули (17) та (18) до іншого вигляду

$$\begin{cases} m_{21} = \dfrac{\sigma_2}{\sigma_1}(m_{11} - m_1) + m_2; \\ m_{22} = \dfrac{\sigma_2}{\sigma_1}(m_{12} - m_1) + m_2; \\ \sigma_{21} = \dfrac{\sigma_2}{\sigma_1}\sigma_{11}; \\ \sigma_{22} = \dfrac{\sigma_2}{\sigma_1}\sigma_{12}. \end{cases} \quad (26)$$

Розглянемо два часткових випадки. Нехай є випадкова величина, що описується щільністю ймовірностей (4), з довільними значеннями математичного очікування m_1 та дисперсії σ_1^2. Отримаємо параметри розподілу (4), який буде описувати випадкову величину того ж типу але з нульовим математичним очікуванням і одиничною дисперсією. Згідно (26) отримуємо нові значення параметрів суміші:

$$\begin{cases} m_{21} = \dfrac{(m_{11} - m_1)}{\sigma_1}; \\ m_{22} = \dfrac{(m_{12} - m_1)}{\sigma_1}; \\ \sigma_{21} = \dfrac{\sigma_{11}}{\sigma_1}; \\ \sigma_{22} = \dfrac{\sigma_{12}}{\sigma_1}. \end{cases} \quad (27)$$

З іншого боку, в разі коли із стандартної випадкової величини необхідно отримати випадкову величину того ж типу, але з довільними значеннями математичного сподівання і дисперсії згідно (26) отримуємо формули

$$\begin{cases} m_{21} = \sigma_2 m_{11} + m_2; \\ m_{22} = \sigma_2 m_{12} + m_2; \\ \sigma_{21} = \sigma_2 \sigma_{11}; \\ \sigma_{22} = \sigma_2 \sigma_{12}. \end{cases}$$

Отримаємо за допомогою формул (27) параметри розподілу стандартної випадкової величини того ж типу що і випадкові величини, щільності ймовірностей яких представлені на рис. 3, 4:

$d_1 = 0{,}3$; $d_2 = 0{,}7$; $\sigma_1 = 0{,}37$; $\sigma_2 = 1{,}1$;

$m_1 = -0{,}51$; $m_2 = 0{,}22$. (28)

На рис. 5 представлена суміш розподілів з параметрами (28), і, для порівняння, щільність ймовірностей стандартного нормального розподілу $N(x)$.

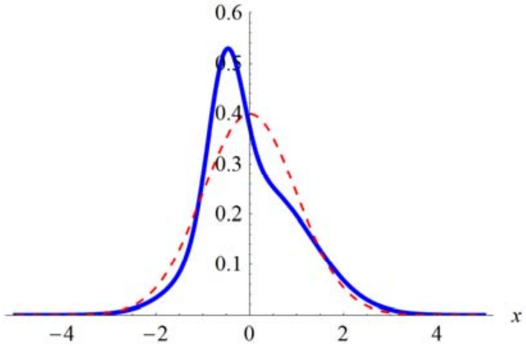

Рис. 5. Щільність ймовірностей суміші $p(x)$ (▬) і нормальної стандартної величини $N(x)$ (- -)

Висновки

В даній роботі отримано співвідношення, що визначає функцію розподілу гаусівских сумішей одного типу та отримано вирази, що визначають залежність між параметрами сумішей розподілів одного типу. Результат використання цих виразів наочно демонструє приклад, який показав, що незважаючи на відмінність в параметрах двох розподілів вони належать до одного типу і відрізняються тільки математичними сподіваннями і дисперсіями.

Виведені співвідношення переходу від стандартної випадкової величини з розподілом (3) до випадкової величиною з довільними математичним очікуванням і дисперсією, і навпаки дозволяють уніфікувати аналіз властивостей сумішей розподілів.

У подальшому результати роботи будуть використані при розв’язанні задач класифікації сумішей розподілу а також для отримання ймовірнісних характеристик оцінок параметрів сумішей розподілів.

Література

1. *Величкин А.И.* Передача аналоговых сообщений по цифровым каналам связи. – М.: Радио и связь, 1983. – 240 с.
2. *Горовецкая Т.А.*, Красильников А.И., Чан Хыу Дат. Модели и законы распределения флуктуационных сигналов // Электроника и связь. – 2000. – № 9. – С. 5–14.
3. *K.N. Plataniotis* and D. Hatzinakos. Gaussian Mixtures and Their Applications to Signal Processing Handbook. Editor Stregios Stregiopolus. – Boca Raton: CRC Press LLC, 2001. – P. 47.
4. *Лукач. Е.* Характеристические функции / Пер. с англ. – М.: Наука, 1979. – 424 с.
5. *Феллер В.* Введение в теорию вероятностей и ее приложения. В 2 томах. Т. 1 / Пер. с англ. – М.: Мир, 1984. – 528 с.
6. *Красильников А.И.*, Пилипенко К.П. Применение двухкомпонентной гауссовской смеси для идентификации одновершинных симметричных плотностей вероятностей // Электроника и связь. – 2008. – № 5(46). – С. 20–29.
7. *Красильников А.И.*, Пилипенко К.П. Одновершинная двухкомпонентная гауссовская смесь. Коэффициент эксцесса // Электроника и связь. – 2007. – № 2(37). – С. 32–38.
8. *Тихонов В.И.* Статистическая радиотехника. – 2-е изд., перераб и доп.– М.: Радио и связь, 1982. – 624с.

Национальный технический университет Украины «Киевский политехнический институт»

Поступила в редакцию 18 июня 2012 г.