ƯỚC LƯỢNG BAYES CHO TỶ LỆ TRỘN TRONG PHÂN LOẠI VÀ NHẬN DẠNG HAI TỔNG THỂ

Võ Văn Tài(1), Phạm Gia Thụ(2), Tô Anh Dũng(3)
(1) Trường Đại học Cần Thơ
(2)Trường Đại học Moncton, Canada
(3)Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM
(Bài nhận ngày 15 tháng 04 năm 2007, hoàn chỉnh sửa chữa ngày 10 tháng 09 năm 2007)

***TÓM TẮT**: Bài báo trình bày một bài toán phân loại và nhận dạng hai tổng thể H_1 và H_2 bằng phương pháp Bayes, đó là xây dựng hàm mật độ xác suất hậu nghiệm cho v, tỷ lệ trộn của H_1 trong H_3 (phần trộn của H_1 và H_2) dựa trên phân phối tiên nghiệm của v chặt cụt trên khoảng (0,1) và sử dụng quan sát lấy từ H_3. Các trường hợp v có phân phối tiên nghiệm beta, mũ và chuẩn được xem xét chi tiết.*
***Từ khóa**: Tiên nghiệm, hậu nghiệm, phân phối mũ, beta, chuẩn.*

1. GIỚI THIỆU

Trong thực tế có nhiều vấn đề đòi hỏi chúng ta phải giải quyết bài toán phân loại và nhận dạng hai tổng thể H_1 và H_2. Có nhiều cách khác nhau để giải quyết bài toán phân loại này. Một phương pháp phân loại có nhiều ưu điểm dựa trên hàm mật độ xác suất của hai tổng thể đó là phương pháp Bayes. Trong phân loại này người ta quan tâm đến tổng thể H_3 chứa những phần tử chung của H_1 và H_2, kết hợp từ mỗi tổng thể với tỷ lệ nào đó.

Giả sử trên H_1 và H_2 ta quan sát biến ngẫu nhiên X, ký hiệu $f_1(x)$, $f_2(x)$ là hàm mật độ xác suất tương ứng của X trên hai tổng thể, và gọi v là **tỷ lệ trộn** của những phần tử của H_1 trong H_3 $(0<v<1)$, khi đó hàm mật độ xác suất của X trên H_3 có dạng $g(x)=v\,f_1(x)+(1-v)\,f_2(x)$. Tham số v thường không được biết một cách chính xác, vì vậy quan tâm của chúng tôi ở đây là tìm cách ước lượng v.

Ước lượng này đã được nghiên cứu bởi Everitt (1985), McLachlan và Basford (1988) [2] bằng phương pháp cực đại tỷ số hợp lý và phương pháp moment. Trước đó James (1978) dựa trên thực tế để ước lượng v. Một phương pháp ước lượng đáng chú ý phải kể đến của nhóm tác giả Pham-Gia, N. Turkkan và A. Bekker (2005) [4]. Họ đã sử dụng phương pháp Bayes để ước lượng cho v với giả thiết v có luật phân phối xác suất tiên nghiệm cụ thể beta và sử dụng phân tích nhận dạng dựa trên mẫu loại Bernoulli lấy từ H_3 để xác định số phần tử thuộc H_1 nằm trong H_3.(điều này có thể làm được vì H_1 và H_2 đã được xác định), để từ đó tìm phân phối xác suất hậu nghiệm cho v.

Trong bài viết này chúng tôi tiếp tục phát triển ý tưởng trên với giả thiết v có phân phối tiên nghiệm **bất kỳ** nào đó $f_{prior}(v)$ để tìm hàm mật độ xác suất hậu nghiệm cho v. Vấn đề trên cũng được chúng tôi xem xét cụ thể khi $f_{prior}(v)$ là hàm mật độ xác suất mũ và chuẩn chặt cụt trên (0, 1). Trong [4] xét phân phối beta chuẩn trên (0, 1) và đó là điều đương nhiên vì phân phối này xác định trên (0, 1). Tuy nhiên với phân phối chuẩn thì miền xác định là cả trục số. Để có thể có được tỷ lệ, mà tỷ lệ phải nằm trong khoảng (0, 1) vì vậy chúng tôi đưa ra ý tưởng mới là "chặt cụt" phân phối chuẩn trên khoảng (0, 1). Ngoài ra, hàm mật độ xác suất hậu nghiệm có dạng đóng của v được xác định, đó là hình thức tốt nhất giúp ước lượng v, điều này có ý nghĩa rất quan trọng trong phân loại và nhận dạng hai tổng thể. Bài báo cũng đưa ra chương trình tính toán soạn trên phần mềm MAPLE.

2.HÀM MẬT ĐỘ HẬU NGHIỆM DẠNG ĐÓNG CỦA *V* TRONG THỐNG KÊ BAYES VỚI MẪU LOẠI BERNOULLI

2.1 Định lý về hàm mật độ xác suất hậu nghiệm của *v*

Định lý 1: *Gọi v là tỷ lệ trộn của tổng thể* H_1 *trong tổng thể* H_3, τ *và* δ *lần lượt là xác suất sai lầm khi phân loại giữa* H_1 *và* H_2. *Nếu v có phân phối tiên nghiệm* $f_{prior}(v)$ *và với n quan sát từ* H_3, *trong số đó có j quan sát thuộc* H_1 *thì v sẽ có hàm mật độ xác suất hậu nghiệm là*

$$\varphi^{(n,j)}(v) = \frac{f_{prior}(v)\left[1-Av\right]^j\left[1-Bv\right]^{n-j}}{L(n,j)} \tag{1}$$

trong đó,

$$A=(\tau+\delta-1)/\delta; \quad B=(1-\tau-\delta)/(1-\delta); \quad L(n,j)=\int_0^1 f_{prior}(v)[1-Av]^j[1-Bv]^{n-j}\,dv$$

Chứng minh.

Khi v là hằng số, trong phân loại của H_1 và H_2 ta có xác suất phân loại sai lầm:

$\tau = P(H_2 \mid H_1)$: Xác suất phân loại một phần tử vào H_2 khi thật sự nó thuộc H_1.

$\delta = P(H_1 \mid H_2)$: Xác suất phân loại một phần tử vào H_1 khi thật sự nó thuộc H_2.

Khi lấy một mẫu từ H_3 thì xác suất chọn được một phần tử của H_1 là θ có dạng

$$\begin{aligned}\theta &= P(H_1)P(H_1 \mid H_1) + P(H_2)P(H_1 \mid H_2)\\ &= v(1-\tau)+(1-v)\delta = \delta+(1-\delta-\tau)v\end{aligned}$$

trong đó, $P(H_1 \mid H_1)$ có nghĩa là xác suất phân loại một phần tử của H_1 vào đúng H_1.

Hàm hợp lý khi lấy một mẫu gồm *n* phần tử từ tổng thể H_3 có *j* phần tử thuộc H_1 là

$$\begin{aligned}f_{likelihood}(v) &= \theta^j(1-\theta)^{n-j}\\ &= [\delta+(1-\delta-\tau)v]^j[1-(\delta+(1-\delta-\tau)v]^{n-j}\\ &= \delta^j(1-Av)^j(1-\delta)^{n-j}(1-Bv)^{n-j}\end{aligned}$$

Nếu v không biết, ta xem v là một biến ngẫu nhiên với hàm mật độ xác suất tiên nghiệm là $f_{prior}(v)$.

Đặt $K=\int_0^1 f_{prior}(v)f_{likelihood}(v)\,dv = \delta^j(1-\delta)^{n-j}L(n,j)$, khi đó, hàm mật độ hậu nghiệm của v sẽ có dạng:

$$\varphi^{(n,j)}(v) = \frac{f_{prior}(v)f_{likelihood}(v)}{K} = \frac{f_{prior}(v)\left[1-Av\right]^j\left[1-Bv\right]^{n-j}}{L(n,j)} \qquad \square$$

2.2 Một số trường hợp riêng của định lý

Hệ quả 1: *Khi v có phân phối tiên nghiệm Beta* $(v;\alpha,\beta)$ *với* $\alpha,\beta>0$

a) *v có hàm mật độ xác suất hậu nghiệm*

$$\varphi^{(n,j)}(v) = Beta(v;\alpha,\beta).[1-Av]^j[1-Bv]^{n-j} / P_0^{(n,j)}(A,B) \quad (2)$$

trong đó $P_0^{(n,j)}(A,B) = \int_0^1 v^{\alpha-1}(1-v)^{\beta-1}(1-Av)^j(1-Bv)^{n-j}\,dv$ *mà ta có thể tính theo định lý Picard* (xem [4]).

b) *Trung bình hậu nghiệm của v là*

$$\mu^{(n,j)}(v) = \mu_{prior}(v).\frac{P_1^{(n,j)}(A,B)}{P_0^{(n,j)}(A,B)} \quad (3)$$

c) *Phương sai hậu nghiệm của v là*

$$Var^{(n,j)}(v) = \frac{Var_{prior}(v)}{P_0^2(A,B)(1-\mu_{prior}(v))}\Big\{(\alpha+1)P_0^{(n,j)}(A,B)P_2^{(n,j)}(A,B) -$$

$$-(\alpha+\mu_{prior}(v))\left(P_1^{(n,j)}(A,B)\right)^2\Big\} \quad (4)$$

Đây là định lý đã được tác giả T. Pham-Gia trình bày trong [4].

Hệ quả 2: *Khi v có phân phối tiên nghiệm mũ chặt cụt trên (0,1) với tham số b > 0*

a) *v có phân phối hậu nghiệm*:

$$\varphi^{(n,j)}(v) = \begin{cases} \dfrac{Exp(b)(1-Av)^j(1-Bv)^{n-j}}{I(n,j)}, & v\in(0,1) \\ 0, \; v\notin(0,1) \end{cases} \quad (4)$$

trong đó, $I(n,j) = \int_0^1 be^{-bv}(1-Av)^j(1-Bv)^{n-j}\,dv$ *mà ta có thể dùng tích phân truy hồi để tính* (Xem phụ lục I*)*.

b) *Trung bình hậu nghiệm của v là*

$$\mu^{(n,j)}(v) = \frac{1}{(B-A)I(n,j)}[I(n+1,j+1) - I(n+1,j)] \quad (5)$$

c) *Phương sai hậu nghiệm của v là*

$$Var^{(n,j)}(v) = \frac{1}{AB.I(n,j)}I(n+2,j+2) + \frac{A+B}{AB(B-A)I(n,j)}[I(n+1,j+1) - \quad (6)$$

$$-I(n+1,j)] - \frac{1}{AB} - \left[\mu^{(n,j)}(v)\right]^2$$

Chứng minh.

a) Thay $f_{prior}(v)$ bằng phân phối mũ chặt cụt trên (0,1) $f_{prior}(v) = \dfrac{be^{-bv}}{1-e^{-b}}$ vào (1), qua một số tính toán sơ cấp ta có điều phải chứng minh.

b) Vì $v = \dfrac{1}{B-A}[(1-Av)-(1-Bv)]$ nên:

$$\mu^{(n,j)}(v) = \int_0^1 v.\varphi^{(n,j)}(v)dv$$

$$= \frac{1}{(B-A)I(n,j)}\int_0^1 be^{-bv}(1-Av)^j(1-Bv)^{n-j}\left[(1-Av)-(1-Bv)\right]dv$$

$$= \frac{1}{(B-A)I(n,j)}\left[I(n+1,j+1)-I(n+1,j)\right]$$

c) Tương tự bằng cách thế

$$v^2 = \frac{1}{AB}\left[(1-Av)(1-Bv)+\frac{A+B}{B-A}(1-Av)+\frac{A+B}{A-B}(1-Bv)-1\right]$$

vào biểu thức $Var^{(n,j)}(v) = \int_0^v v^2\varphi^{(n,j)}(v)dv - \left[\mu^{(n,j)}(v)\right]^2$ và qua một số bước tính sơ cấp ta có kết quả (6). □

Hệ quả 3: *v có phân phối tiên nghiệm chuẩn chặt cụt trên (0,1) với hai tham số* μ,σ

a) *v có phân phối hậu nghiệm là*

$$\varphi^{(n,j)}(v) = \begin{cases} \dfrac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}.(1-Av)^j(1-Bv)^{n-j}}{L(n,j)}, & v\in(0,1) \\ 0, \quad v\notin(0,1) \end{cases} \tag{7}$$

Trong đó $L(n,j) = \int_0^1 e^{\frac{-(x-\mu)^2}{2\sigma^2}}.(1-Av)^j(1-Bv)^{n-j}dv$ *mà ta có thể dùng tích phân truy hồi để tính* (Xem phụ lục II).

b) *Trung bình hậu nghiệm của v là*

$$\mu^{(n,j)}(v) = \frac{1}{(B-A)L(n,j)}[L(n+1,j+1)-L(n+1,j)] \tag{8}$$

c) *Phương sai hậu nghiệm của v là*

$$\begin{aligned} Var^{(n,j)}(v) = {} & \frac{1}{AB.L(n,j)}L(n+2,j+1)+\frac{A+B}{AB(B-A)L(n,j)}\left[L(n+1,j+1)-\right. \\ & \left.-L(n+1,j)\right]-\frac{1}{AB}-\left[\mu^{(n,j)}(v)\right]^2 \end{aligned} \tag{9}$$

Chứng minh.

Hoàn toàn tương tự như chứng minh hệ quả 2, chỉ thay *I(n,j)* bởi *L(n,j)*. □

3. VÍ DỤ SỐ

3.1.Bài toán.

Giả sử H_1 và H_2 liên kết tạo ra tổng thể H_3 mà trong đó H_1 chiếm tỷ lệ là *v*. Giá trị chính xác của *v* chưa biết. Giả sử *v* là biến ngẫu nhiên có phân phối tiên nghiệm tuân theo luật chuẩn chặt cụt trên (0, 1) hay phân phối mũ chặt cụt trên (0, 1), và khi lấy một mẫu gồm 20 quan sát từ H_3 có 5 quan sát thuộc H_1. Cần xác định hàm mật độ xác suất hậu nghiệm cho v.

3.2.Giải.

Nếu trên hai tổng thể H_1 và H_2 ta quan sát biến ngẫu nhiên X_1 và X_2 lần lượt có phân phối chuẩn $X_1 \sim N(5, 9^2)$, $X_2 \sim N(18, 6^2)$ thì phương trình $f_1(x) = f_2(x)$ có hai nghiệm x_1=11.198 và x_2 = 45.602. Vì vậy trong phân tích nhận dạng bằng phương pháp Bayes nếu kết quả quan sát là $11.198 \le x \le 45.602$ thì quan sát đó được xếp vào H_1, ngược lại xếp vào H_2. Trong phân tích nhận dạng này hai xác suất sai lầm được tính cụ thể như sau:

$$\tau = \int_{11.198}^{45.602} f_1(x)dx = 0.2455$$

$$\delta = \int_{-\infty}^{11.198} f_2(x)dx + \int_{45.602}^{+\infty} f_1(x)dx = 0.1285$$

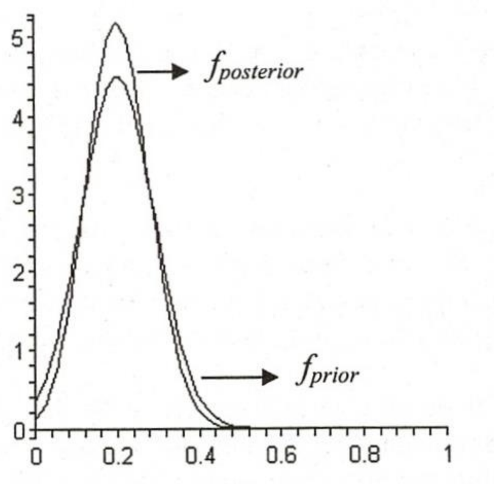
Với τ và δ trên thì $A = -4.872$, $B = 0.718$.

Nếu v có phân phối tiên nghiệm chuẩn $N(0.2;\ 0.09^2)$ chặt cụt trên (0,1), thì hàm mật độ xác suất tiên nghiệm của nó là:

$$f_{prior}(v) = \frac{5.6295\sqrt{2}e^{-61.7284(v-0.2)^2}}{\sqrt{\pi}} = 4.4947.e^{-61.7284(v-0.2)^2}$$

Khi đó, vì $L(20, 5) = 0.56838$, nên hàm mật độ hậu nghiệm của v theo hệ quả 3 sẽ là:

$$f_{posteroor}(v) = 1.7594e^{-61.7284(v-0.2)^2}(1+4.872v)^5(1-0.718v)^{15}$$



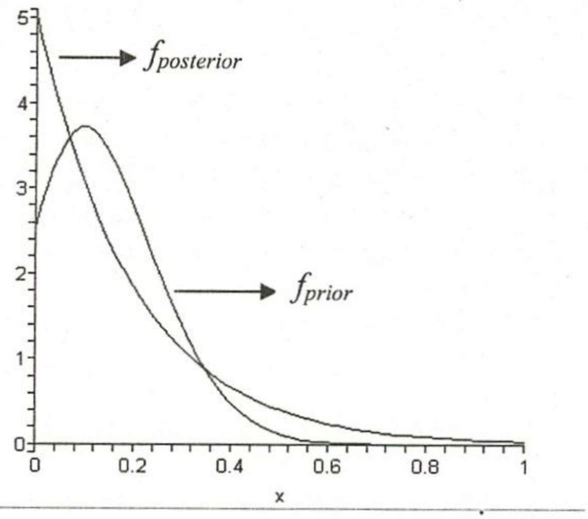
Hình 1: Đồ thị của hàm mật độ tiên nghiệm, hậu nghiệm chuẩn $N(0.2, 0.09^2)$ của v chặt cụt trên (0,1)

Nếu v có phân phối tiên nghiệm mũ, $v \sim Exp(5)$ chặt cụt trên (0,1) thì hàm mật độ xác suất tiên nghiệm của nó là:

$$f_{prior}(v) = 5.034e^{-5v}$$

Khi đó, do $I(20,5) = 1.9377$ nên hàm mật độ hậu nghiệm của v theo hệ quả 2 sẽ là

$$f_{posterior}(v) = 2.5804e^{-5v}(1+4.872v)^5(1-0.718v)^{15}$$



Hình 2: Đồ thị của hàm mật độ tiên nghiệm, hậu nghiệm mũ Exp*(5)* của v chặt cụt trên (0,1)

Các tham số trung bình, phương sai của phân phối tiên nghiệm mũ và chuẩn chặt cụt trên (0,1) và phân phối hậu nghiệm của v tính bởi (5), (6), (8) và (9) cho ở bảng sau:

	μ_{prior}	μ_{pos}	Var_{prior}	Var_{pos}
Phân phối mũ	0.20308	0.20197	0.00747	0.00581
Phân phối chuẩn	0.19322	0.16383	0.03317	0.01241

Như vậy phân phối hậu nghiệm của v trong cả hai trường hợp đều có phương sai nhỏ hơn đáng kể so với phân phối tiên nghiệm của nó, nghĩa là phân phối hậu nghiệm của v gọn hơn, cho phép chúng ta đánh giá được v một cách chính xác hơn.

4. KẾT LUẬN

Bằng phương pháp phân loại và nhận dạng Bayes, bài báo đã đưa ra được công thức chung để xác định hàm mật độ xác suất của v, và xét cụ thể khi v cĩ phân phối chuẩn và mũ chặt cụt trên (0, 1). Chương trình tính toán bằng phần mềm **MAPLE** của bài báo có thể rất hữu ích trong thực tế. Sắp tới chúng tôi sẽ viết lại chương trình thân thiện hơn với người dùng dưới dạng cửa sổ.

Việc ước lượng tham số v có một lợi ích rất lớn khi phân loại và nhận dạng hai tổng thể bằng phương pháp Bayes, đặc biệt là trong việc đánh giá sai số trong phân loại.

Vấn đề trên sẽ phức tạp hơn rất nhiều khi phân loại hơn hai tổng thể, chúng tôi sẽ trình bày trong các bài báo sau.

Phụ lục I

Tính $I(n,k) = \int_0^1 be^{-bv}(1-Av)^k(1-Bv)^{n-k}\,dv$

Khi $n = k = 0$ thì $I(0,0) = 1- e^{-b}$

Khi $k = 0.\ n > 0$ thì $I(n,0) = \int_0^1 be^{-bv}(1-Bv)^n\,dv = 1 - e^{-b}(1-B)^n - \dfrac{nB}{b}I(n-1,0)$

Khi $n = k > 0$ thì $I(n,n) = \int_0^1 be^{-bv}(1-Av)^n\,dv = 1 - e^{-b}(1-A)^n - \dfrac{nA}{b}I(n-1,n-1)$

Khi $n > k > 0$, đặt $U = (1-Av)^k(1-Bv)^{n-k}$ ta có $dV = be^{-bv}dv$ và
$dU = - kA(1-Av)^{k-1}(1-Bv)^{n-k} - (n-k)B(1-Av)^k(1-Bv)^{n-k-1}dv;\ \ v = - e^{-bv}$

Khi đó:

$$\begin{aligned} I(n,k) &= [UV]_0^1 - \int_0^1 VdU \\ &= \left[-e^{-bv}(1-Av)^k(1-Bv)^{n-k}\right]_0^1 - \frac{kA}{b}I(n-1,k-1) - \frac{(n-k)B}{b}I(n-1,k) \\ &= 1- e^{-b}(1-A)^k(1-B)^{b-k} - \frac{kA}{b}I(n-1,k-1) - \frac{(n-k)B}{b}I(n-1,k) \end{aligned}$$

Tích phân truy hồi này dẫn đến các tích phân $I(0,0)$, $I(p,0)$ và$I(p,p)$ có thể tính được ở trên. Như vậy bằng việc sử dụng các vòng lặp của tích phân truy hồi, $I(n,k)$ được tính một cách tổng quát.

Và để tính $I(p,q)$ với tham số b chúng tôi lập trình tính toán bằng phần mềm **MAPLE**:

```
tp1:=proc(p,q::nonnegint);
if p=0 then
evalf[15](1-exp(-b));
else if p>0 and q =0 then
evalf[15](1-exp(-b)*(1-B)^p-((p*B)/b)*tp1(p-1,q));
fi; fi; end:

tp2:=proc(p,q::nonnegint);
if p=0 and q=0 then
evalf[15](1-exp(-b));
else if p=q then
evalf[15](1-exp(-b)*(1-A)^p-((p*A)/b)*tp2(p-1,q-1));
fi;fi;end:

tp3:=proc(p,q::nonnegint);
if q=0 and p=0 then evalf[15](1-exp(-b));
else if q=0 and p>0 then tp1(p,q);
else if p=q and p>0 and q>0 then tp2(p,q);
else if q>0 and p>0 and p>q then
evalf[15](1-exp(-b)*((1-A)^q)*(1-B)^{(p-q)}-((q*A)/b)*tp3(p-1,q-1)-(((p-q)*B)/b)*tp3(p-1,q));
fi;fi;fi;fi;end:
```

Phụ lục II

Tính $L(n,k) = \int_0^1 e^{-\frac{(v-\mu)^2}{2\sigma^2}} (1-Av)^k (1-Bv)^{n-k}\, dv$

$$L(n,k) = \int_0^1 e^{-\frac{(v-\mu)^2}{2}} (1-Av)^k (1-Bv)^{n-k}\, dv$$

$$= -A\sigma^2 \int_0^1 \frac{v-\mu}{\sigma^2} e^{-\frac{(v-\mu)^2}{2\sigma^2}} (1-Av)^{k-1}(1-Bv)^{n-k}\, dv$$

$$+(1-A\mu)\int_0^1 e^{-\frac{(v-\mu)^2}{2\sigma^2}} (1-Av)^{k-1}(1-Bv)^{n-k}\, dv$$

$$= A\sigma^2 Q + (1-A\mu)L(n-1,k-1)$$

Tính $Q = -\int_0^1 \frac{v-\mu}{\sigma^2} e^{-\frac{(v-\mu)^2}{2\sigma^2}} (1-Av)^{k-1}(1-Bv)^{n-k}\, dv$

Đặt $U = (1-Av)^{k-1}(1-Bv)^{n-k}$, $dV = -\frac{v-\mu}{\sigma^2} e^{-\frac{(v-\mu)^2}{2\sigma^2}}$

$$dU = [-(k-1)A(1-Av)^{k-2}(1-Bv)^{n-k} - (n-k)B(1-Av)^{k-1}(1-Bv)^{n-k-1}]dv;\quad V = e^{-\frac{(v-\mu)^2}{2\sigma^2}}$$

Khi đó $Q = [UV]_0^1 - \int_0^1 V dU$

$$= \left[e^{-\frac{(v-\mu)^2}{2\sigma^2}} (1-Av)^{k-1}(1-Bv)^{n-k} \right]_0^1 + (k-1)AL(n-2,k-2) + (n-k)BL(n-2,k-1)$$

$$= e^{-\frac{(1-\mu)^2}{2\sigma^2}} (1-A)^{k-1}(1-B)^{n-k} - e^{-\frac{\mu^2}{2\sigma^2}} + (k-1)AL(n-2,k-2) + (n-k)BL(n-2,k-1)$$

Vậy $L(n,k) = A\sigma^2 \left[e^{-\frac{(1-\mu)^2}{2\sigma^2}} (1-A)^{k-1}(1-B)^{n-k} - e^{-\frac{\mu^2}{2\sigma^2}} \right] +$

$$+ (k-1)A^2\sigma^2 L(n-2,k-2) + (n-k)AB\sigma^2 L(n-2,k-1) + (1-A\mu)L(n-1,k-1).$$

Tích phân truy hồi này sẽ dẫn đến các tích phân $L(p,p)$, $L(p,0)$ và $L(p+1,p)$.

$L_A(p) = L(p,p)$ được tính tổng quát bằng cách thay thế $n = k = p$, cụ thể:

$$L_A(p) = A\sigma^2 \left[e^{-\frac{(1-\mu)^2}{2\sigma^2}} (1-A)^{p-1} - e^{-\frac{\mu^2}{2\sigma^2}} \right] + (p-1)A^2\sigma^2 L(p-2,p-2)$$

$$+ (1-A\mu)L(p-1,p-1)$$

Trong đó $L_A(0) = \int_0^1 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)$.

$L(p,0) = L_B(p)$.

$$L(p+1,p) = B\sigma^2\left[e^{-\frac{(1-\mu)^2}{2\sigma^2}}(1-A)^p - e^{-\frac{\mu^2}{2\sigma^2}}\right] + pAB\sigma^2 L_A(p-1) + (1-B\mu)L_A(p).$$

Bằng việc sử dụng các vòng lặp ta có thể tính được tích phân $L(n,k)$.

Chương trình tính $L(p,q)$ với trung bình ***m*** và độ lệch chuẩn ***a*** bằng phần mềm **MAPLE** như sau:

```
tp1:= proc(p,q::nonnegint);
if p=0 and q=0 then
evalf[15](int(exp(-(x-m)^2/2*a^2),x=0..1));
else if p=1 and q=0 then
evalf[15] (int(exp(-(x-m)^2/2*a^2)*(1-B*x),x=0..1));
else if p>1 and q=0 then
evalf[15](B*a^2*(exp(-(1-m)^2/2*a^2)*(1-B)^{(p-1)}-exp(-m^2/2*a^2))
    +(p-1)*B^2*a^2*tp1(p-2,0)+(1-B*m)*tp1(p-1,0));
fi;fi;fi;end:

tp2:=proc(p,q::nonnegint);
if q=0 then tp2(p,q)
else if p =1 and q =1 then
evalf[15](int(exp(-(x-m)^2/2*a^2)*(1-A*x),x=0..1));
else if p=2 and q=1 then
evalf[15](int(exp(-(x-m)^2/2*a^2)*(1-A*x)*(1-B*x),x=0..1));
else if p >2 and q=1 then
evalf[15](A*a^2*(exp(-(1-m)^2/2*a^2)*(1-B)^{(p-1)}-exp(-m^2/2*a^2))
   + (p-1)*A*B*a^2*tp2(p-2,0)+(1-A*m)*tp2(p-1,0));
fi;fi;fi;fi;  end:

tp3:=proc(p,q::nonnegint);
if q=0 then tp1(p,q);
else if q=1 then tp2(p,q);
else if q>1  then
evalf[15](A*a^2*(exp(-(1-m)^2/2*a^2)*(1-A)^{(q-1)}*(1-B)^{(p-q)}-exp(-m^2/2*a^2))+(q-
1)*A^2*a^2*tp3(p-2,q-2)+(p-q)*A*B*a^2*tp3(p-2,q-1)+(1-A*m)*tp3(p-1,q-1));
fi;fi;fi; end:
```

BAYESIAN ESTIMATION FOR THE MIXING PROPORTION IN CLASSIFICATION AND DISCRIMINANT WITH TWO POPULATIONS

Vo Van Tai[(1)], Pham Gia Thu[(2)], To Anh Dung[(3)]

[(1)] Can tho University

[(2)] Moncton University, Canada

[(3)] University of Natural Sciences, VNU-HCM

ABSTRACT: *The article presents a problem in classification with two populations H_1 and H_2 by Bayesian method, which is building posterior probability distribution function for v,* *proportion of H_1 in H_3 (the mixing of H_1 and H_2) using its truncated (0,1) prior distribution and observations from H_3. The cases when v has prior beta, normal and exponential distributions are studied completely.*

Keywords: *Prior distribution, posterior distribution, exponential, beta, normal.*

TÀI LIỆU THAM KHẢO

[1]. Andrew R. Webb. *Statistical Pattern Recognition*. John Wiley, London, (1999).

[2]. Maclachlan G., Basford K. *Mixture models*. Marcel Dekker, New York, (1988).

[3]. Morris H.Degroot. *Probability and Statistics*. Addison-Wesley, United State, (1986).

[4]. Pham-Gia T., Turkkan N., Bekker A. *Bayesian Analysis in the L^1 – Norm of the Mixing Proportion using Discriminant Analysis*. Metrika, (2005).