المجلة العراقية للعلوم الاحصائية (12) 2007
ص ص [114–124]

التقدير والتنبؤ في التوزيع الطبيعي المبتور المختلط

طالب شريف جليل* هيفاء عبد الجواد سعيد**

المستخلص

يعنى هذا البحث بتقدير معلمات التوزيع الطبيعي المختلط بمركبتين مبتورتين من طرفين وذلك باستخدام أسلوب بيز. ويستخدم إحدى الطرائق العددية في إيجاد التوزيع اللاحق المشترك لتباين مركبتي التوزيع المختلط . كما عني أيضا بإيجاد التوزيع التنبؤي للمشاهدة المستقبلية وتبين بأنه توزيع مختلط موزون بعائدية المشاهدة المستقبلية الى كل من مركبتي التوزيع الطبيعي المبتور المختلط.

Estimation and Prediction in Mixture Truncated Normal Distribution

Abstract

This paper is concerned with the estimation of the parameters of the mixture normal distribution with truncated components using Bayesian approach. One of numerical methods is used to obtain the joint posterior distribution of the variances of the two components, Predictive distribution of the future observation, is also obtained which is a mixture of two components. Each component is weighted by the belongance ratio of future observation to each component of mixture truncated normal distribution.

1.المقدمة

للتوزيعات المبتورة المختلطة أهمية كبيرة في اختبار الأعمار الزمنية للأجهزة الكهربائية وقد لاقت هذه التوزيعات اهتماماً كبيراً من قبل بعض الباحثين الاحصائيين.فقد درس (David & Johnson(1952))[5] المقدرات بطريقة الترجيح

---

*استاذ مساعد/ كلية علوم الحاسبات والرياضيات/قسم الإحصاء والمعلوماتية
** مدرس/ كلية علوم الحاسبات والرياضيات/قسم الإحصاء والمعلوماتية.

تاريخ التسلم : 4 /2007/3 ———— تاريخ القبول : 23 /2007/5

الأعظم مع العينات المبتورة وعني ((Cohen(1960)[2] بمسألة التقدير عند فقدان الفئتين الصفر والواحد، وهؤلاء الاحصائيون عاملوا (treated) الفئة الصفرية على أنها مفقودة (missing) على الرغم من أنها غير مفقودة بل مخلوط مع توزيع آخر.واستخدم ((Cohen (1966)[3] دالة كثافة احتمال مختلطة تحتوي على مركبة ذات توزيع صفري مع مركبة ذات توزيع بواسون المبتور وقدر معلماته بطريقة الترجيح الأعظم. وعنيت (سعيد (2005))[1] بمسألة تقدير المعلمات للتوزيعات المختلطة المبتورة والمبتورة المختلطة بطريقة الترجيح الأعظم وبأسلوب بيز لمجموعة مختلفة من التوزيعات وعني أيضا بالتوزيع التنبؤي للمشاهدة المستقبلية المسحوبة من المجتمع الطبيعي الملوث ذي مركبتين وثلاثة مركبات مبتورة.

تركز اهتمامنا في هذا البحث على التوزيع التنبؤي للمشاهدة المستقبلية المسحوبة من المجتمع الطبيعي الملوث بمركبتين مبتورتين اخذين بنظر الاعتبار أن جميع معلماته غير معلومة باستخدام أسلوب بيز.

يتناول المبحث الثاني توضيحاً للبيانات المعروفة المصدر وغير المعروفة المصدر ويتضمن المبحث الثالث ، استخدام أسلوب بيز في تقدير معلمات التوزيع الطبيعي الملوث بمركبتين مبتورتين في حالة كون جميع معلماته غير معلومة اذ يتم الحصول على مقدرات معلمات كل مركبة من مركباته باستخدام الأسلوب التنبؤي للمشاهدة المستقبلية ويستخدم التكامل العددي في إيجاد التوزيع اللاحق المشترك الحدي لتباين مركبتي التوزيع المختلط واستخدم أيضا مرة أخرى التكامل العددي في إيجاد التوزيع التنبؤي للمشاهدة المستقبلية.

2.تعريف البيانات المعروفة المصدر وغير معروفة المصدر

نفرض من البداية أن المتجه العشوائي $\underline{x}=(x_1,x_2,...,x_d)'$ معرف على الفضاء R^d حيث أن $d\geq 1$ والذي له دالة كثافة أو كتلة احتمال مختلطة تحتوي على k من المركبات ومعرفة على وفق المعادلة الآتية:

$$f(\underline{x}\mid\underline{\psi})=\sum_{j=1}^{k}P_j f_j(\underline{x}\mid\underline{\theta}_j) \qquad \dots (1)$$

حيث أن : P_j يمثل نسبة الخلط بحيث أن $0<P_j<1$ و $\sum_{j=1}^{k}P_j=1$ وان $\underline{\psi}=\left(P_1,P_2,\dots,P_k \quad \theta_1,\dots,\theta_k\right)$ يمثل معلمات التوزيع المختلط و $f_j(\underline{x}\mid\underline{\theta}_j)$ تمثل دالة كثافة أو كتلة احتمال المركبة j وإذا توفرت لبيانات المشاهدة فعلاً $x=(x_1,x_2,\dots,x_n)'$ توجد مقابل كل مشاهدة $\underline{x}_i$ $S(i=1,2,\dots,k)$ متغير عشوائي إضافية غير مشاهدة هي $\underline{Z}_j=(Z_{i1},Z_{i2},\dots,Z_{ik})'$ بحيث تأخذ Z_{il} وقيمة واحد عندما تظهر المشاهدة $\underline{x}_i$ في المركبة $(l=1,2,\dots,k)$ أي أن $(Z_{il}=1)$ والرقم صفر فيما عدا ذلك يعني أن $\underline{Z}_j=(0\ 0\ \dots\ 0,0\ ..0)'$ وتسمى المشاهدة Z_{il} بالمتغير الأصم (Latent Variable) (سعيد (2006))[1]

إن المشاهدات الشرطية $\underline{x}\mid\underline{Z}_i,\psi$ مستقلة عن بعضها البعض كذلك فان المشاهدات $(\underline{Z}_i\mid\underline{\psi})$ مستقلة أيضا عن بعضها الآخر ولها دوال كثافة وكتلة احتمال معرفتين على وفق المعادلتين الآتيتين : (Corduneanu and Bishop,2001)[4]

$$f(x_i\mid\underline{z}_i,\underline{\psi})=\sum_{j=1}^{K}z_{ij}f_j(x_i\mid\theta_i)=\prod_{j=1}^{K}\left[f_i(x_i\mid\theta_i)\right]^{z_{ii}} \qquad \dots (2)$$

$$f(\underline{z}_i\mid\underline{\psi})=\prod_{j=1}^{K}P_j^{z_{ij}} \qquad \dots (3)$$

أما التوزيع المشترك لـــــ $\underline{z}_i$ و x_i فله دالة الكثافة الاحتمالية:

$$f(x_i,z_i\mid\underline{\psi})=f(x_i\mid\underline{z}_i,\underline{\psi})f(\underline{z}_i\mid\underline{\psi})$$

وباستخدام نظرية بيز، يكون التوزيع الشرطي لــ $\underline{z}_i \mid x_i, \underline{\psi}$ بالشكل الآتي :

$$f(x_i, \underline{z}_i \mid \underline{\psi}) = \frac{\prod_{j=1}^{K} P_j^{\,z_{ij}} \prod_{j=1}^{K} (f_i(x_i \mid \theta_i))^{z_{ij}}}{\sum_{j=1}^{K} P_j f_i(x_i \mid \theta_j)}$$

دالة الترجيح للبيانات الكاملة (البيانات المشاهدة والإضافية) بالشكل الآتي:

$$L_c(\underline{\psi} \mid X, Z) = \prod_{i=1}^{n} \prod_{j=1}^{K} (P_j f_j(x_i \mid \theta_j))^{z_{ij}} \qquad \dots (4)$$

نعنى في هذا البحث بالتوزيع الطبيعي المختلط بمــركبتين مبتــورتين مــن طــرفين وباستخدام المعادلة (1)، يكون هذا التوزيع معرفاً على وفق المعادلة الآتية:

$$f(x \mid \underline{\psi}) = P \frac{f_1(x \mid \theta_1, \sigma_1^2)}{F_1(b_1) - F_1(a_1)} + (1-P) \frac{f_2(x \mid \theta_2, \sigma_2^2)}{F_2(b_2) - F_2(a_2)} \qquad \dots (5)$$

حيث أن: $\underline{\Psi} = (P_1\ \theta_1\ \theta_2\ \sigma_1^2\ \sigma_2^2)'$ متجه معلمات التوزيع وغير معلومة.

3.التوزيع اللاحق

نفرض أن التوزيع السابق لنسبة الخلط P و θ_j و σ_j^2لقيم j=1,2 معرفة بالشــكل الآتي: (Coeduneanu and Bishop,2001)[4]

$$p \sim Beta(a_0, b_0)$$

$$\theta_j \sim N(\theta_{0j}, \frac{\sigma_j^2}{k_{0j}}) \qquad ,for\ j = 1,2$$

$$\sigma_j^2 \sim IG(\frac{N_{0j}}{2}, \frac{S_{0j}^2}{2}) \qquad , for\ j = 1,2$$

علماً ان IG: يعني توزيع معكوس كاما.

فإذا سحبت n من المشاهدات من المجتمع المختلط المعرف في العلاقة (5) وبتوفر هذه البيانات المشاهدة مع المتغيرات الإضافية وكذلك الاستفادة من المعلومات السابقة عن المعلمات يكون التوزيع اللاحق المشترك للمعلمات $\left(P_1, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2\right)'$ يعرف على وفق المعادلة الآتية: (سعيد (2005))[1]:-

$$p\left(P, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2 \mid X, Z\right) \propto P^{a_0+n_1-1}\left(i-p\right)^{b_0+n_2-1}$$

$$(F_1(b_1-a_1))^{-n_1}(F_2(b_2-a_2))^{-n_2} \exp-\left[\frac{(k_{01}+n_1)\left(\theta_1-\theta_1^{**}\right)^2}{2\sigma^2}\right]$$

$$(\sigma^2)^{-(\frac{v2}{2}+1)} \ \exp - \frac{s_{T_1}^2}{2\sigma_1^2} \quad \exp - \frac{s_{T_2}^2}{2\sigma_2^2}$$

$$\exp-\left[\frac{(k_{02}+n_2)(\theta_2-\theta_2^{**})^2}{2\sigma_2^2}\right](\sigma_1^2)^{-\left(\frac{v}{2}+1\right)} \qquad \dots (6)$$

حيث أن: $\left(s_{T_j}^2, \theta_j^{**} \quad for\ j = 1,2\right)$ معرفة على وفق المعادلتين الآتيتين:-

$$s_{T_j}^2 = s_{0_j}^2 + \frac{n_j K_{0j}(\bar{x}_j - \theta_{0j})^2}{K_{0j} + n_j} + \sum_{i=1}^{n} z_{ij}(x_i - \bar{x}_j)^2 \qquad \ldots (7)$$

$$\theta_j^{**} = \frac{K_{0j}\theta_{0j} + n_j \bar{x}_j}{K_{0j} + n_j} \qquad \ldots (8)$$

وان: $v_j = v_{0_j} + n_j$

نحصل على مقدرات بيز لمعلمات التوزيع المعرف في المعادلة (5) وذلك بالبحث عن قيم لها تجعل دالة التوزيع اللاحق المشتركة المعرفة في المعادلة (6) اكبر ما يمكـن وهذه المقدرات معرفة بالشكل الآتي: (سعيد (2005))[1]

$$\hat{\theta}_{jB}^{(t+1)} = \theta_j^{**} + \frac{n_j \hat{\sigma}_{jB}^{2(t)}}{\left(K_{0j} + n_j\right)}\left(\hat{f}_{jT(a_j,b_j)}^{(t)}(b_j) - \hat{f}_{jT(a_j,b_j)}^{(t)}(a_j)\right) \qquad \ldots (9)$$

$$for\, j = 1,2$$

$$\hat{\sigma}_{jB}^{2(t+1)} = \left[s_{Tj}^2 + (k_{0j} + n_j)(\hat{\theta}_{jB}^{(t)} - \theta_j^{**})^2\right]\Big[v_j + 2 - n_j((b_j - \hat{\theta}_{jB}^{(t)})\hat{f}_{jT(a_j,b_j)}^{(t)}(b_j)$$

$$- (a_j - \hat{\theta}_{jB}^{(t)})\hat{f}_{jT(a_j,b_j)}^{(t)}(a_j))\Big]^{-1} \; for \;\; j = 1,2 \qquad \ldots \textbf{(10)}$$

بعد حل المعادلتين التكراريتين (9) و (10) وباستخدام قيم ابتدائية للمعلمات ويتم اختيار القيم الابتدائية لــ θ_j مساوية لــ $\frac{a_j + b_j}{2}$ لقيم $j = 1,2$ والقيم الابتدائية لــ σ_j^2 يجب ان تقع في داخل الفضاءات الجزئية لــــ A و B ويمكــن اختيــار القــيم الابتدائية لــ σ_j^2 مساوية لنصف طول كل فضاء جزئـــي ،أي ان $\sigma_1^{2(0)} = \frac{A}{2}$ و

$\sigma_2^{2(0)} = \dfrac{B}{2}$ وعند الاستقرار في قيم المعلمات لمرحلتين تكراريتين متعاقبتين نتوصل إلى مقدرات بيز لمعلمات مركبتي التوزيع الطبيعي المختلط المعرف في المعادلة (5) .وفيما يخص مقدر بيز لنسبة الخلط فهو معرف على وفق المعادلة الآتية:

$$\hat{P}_B = \frac{a_0 + n_1 - 1}{a_0 + b_0 + n - 2}$$

4.التوزيع التنبؤي

إن التوزيع اللاحق المشترك لــ σ_2^2, σ_1^2نحصل عليه من تكامل المعادلة (6) نسبة إلى P, θ_1, θ_2 وبأخذ الشكل الآتي : (سعيد (2005))[1]

$$p(\sigma^2, \sigma_2^2 \mid X, Z) \alpha \int_{a_1}^{b_1}\int_{a_2}^{b_2} p(\theta_1, \theta_2, \sigma^2, \sigma_2^2 \mid X, Z)\, d\theta_1 d\theta_2\, \alpha (\sigma^2)^{-(\frac{\nu_1}{2}+1)}$$

$$\exp - \frac{S_{T1}^2}{2\sigma^2} (\sigma_2^2)^{-(\frac{\nu_2}{2}+1)} \exp - \frac{S_{T2}^2}{2\sigma^2} \int_{a_1}^{b_1}\int_{a_2}^{b_2} S_1(\theta_1, \sigma_1^2) S_2(\theta_2, \sigma_2^2)\, d\theta_1 d\theta_2 \qquad \ldots (11)$$

حيث أن:

$$S_j(\theta_j, \sigma_j^2) = (F_1(b_j - a_j))^{-n_j} \exp - \frac{(k_{0j} + n_j)(\theta_j - \theta_j^{**})^2}{2\sigma_j^2} \qquad for \quad j = 1,2$$

علما ان σ_j معرفة في المعادلة (8).

لانستطيع أن نحصل على التكاملات المعرفة في المعادلة (11)تحليلياً وإنما نحتاج إلى إيجادها باستخدام إحدى الطرائق العددية.ولإجراء تلك التكاملات نقسم فضاء $\theta_1 \in (a_1^*, b_1^*)$ إلى I من الفترات ذات الأطوال المتساوية أي أن :

$$(a_1^*, b_1^*) = (d_1, 2d_1, 3d_1, \ldots, Id_1)$$
$$(a_2^*, b_2^*) = (d_2, 2d_2, 3d_2, \ldots, Jd_2)$$

بذلك يمكن الحصول على التوزيع اللاحق المشترك الحدي لــ σ_2^2,σ_1^2 بتطبيق قاعــدة Simpson's ((Love,Rainville (1969))[6] ويأخذ الشكل الآتي:

$$p(\sigma^2,\sigma_2^2 \mid X,Z)\,\alpha(\sigma_1^2)^{-(\frac{v_1}{2}+1)}(\sigma_2^2)^{-(\frac{v_2}{2}+1)}\exp-\frac{s_{T1}^2}{2\sigma_1^2}\exp-\frac{s_{T2}^2}{2\sigma_2^2}$$
$$\left[\sum_{\theta_1\in(a_1,b_1)}\frac{d_1}{3}.(L_1((i-1)d_1,\sigma_1^2)+4L_1(id_1,\sigma_1^2)+L_1((i+1)d_1,\sigma_1^2)))\right]$$
$$\left[\sum_{\theta_2\in(a_2,b_2)}\frac{d_2}{3}.(L_2((i-1)d_2,\sigma_2^2)+4L_2(id_2,\sigma_2^2)+L_2((i+1)d_2,\sigma_2^2)))\right] \qquad \ldots (13)$$

بعد ذلك نأخذ جزءاً من فضائي σ_2^2,σ_1^2 ثم نقسم كلا من الفضائين الجزئيين إلى فترات ذات أطوال متساوية.

نفرض أن جزءاً من فضاء σ_1^2 هو A وجزءاً من فضاء σ_2^2 هو B بحيث أنــه يــتم تقسيم الفضاء A الى M من الفترات المتساوية و B الى N من الفترات المتساوية،أي ان:

$$A=\left[r,2r,3r,\ldots,M_r\right]$$
$$B=\left[s,2s,3s,\cdots,N_s\right]$$

تطبق قاعدة Simpson's مرة أخرى على المعادلة (13) وباستخدام الفضاءات الجزئية A و B لإيجاد الثابت الطبيعي لدالة التوزيع اللاحق المشترك لــ σ_1^2 و σ_2^2 وبأخــذ الشكل المعرف في المعادلة الآتية: (سعيد (2005))[1]

$$U=\left[\sum_{\substack{\sigma_1^2\in A\\ j\in A}}\sum_{\substack{\theta_1\in(a_1,b_1)\\ \mathrm{i}\in I}}\frac{\mathrm{d}_1r_1}{9}(L_1((i-1)d_1,(j-1)r_1)+4L_1(id_1,jr_1)\right.$$
$$\left.(L_1((i+1)d_1,(j+1)r_1)).(v_1((j-1)r_1)+4v_1(jr_1)+v_1((j+1)r_1))\right]$$

$$.\left[\sum_{\substack{\sigma_2^2\in B\\ j\in B}}\sum_{\substack{\theta_2\in(a_2,b_2)\\ i\in I}}\frac{d_2 r_2}{9}(L_2((i-1)d_2,(j-1)s_1)+4L_2(id_2,js_1)\right.$$

$$\left.\left.+L_2((i+1)d_2,(j+1)s_1)).(v_2((j-1)s_1)+4v_2(js_2)+v_2((j+1)s_2)\right]\right]^{-1}$$

$$\ldots (14)$$

بذلك يكون التوزيع اللاحق المشترك الهامشي(Marginal) النظــري لـــــ σ_1^2,σ_2^2 الهامشي بالشكل الآتي : (سعيد (2005))[1]

$$p(\sigma_1^2,\sigma_2^2\mid X,Z)=Uv_1(\sigma_1^2)v_2(\sigma_2^2)\left[\sum_{\theta_1\in(a_1,b_1)}\frac{d_1}{3}.(L_1((i-1)d_1,\sigma_1^2)\right.$$

$$\left.+4L_1(id_1,\sigma_1^2)+L_1((i+1)d_1,\sigma_1^2)\right]\left[\sum_{\theta_2\in(a_2,b_2)}\frac{d_2}{3}(L_2((i-1)d_2,\sigma_2^2\right.$$

$$\left.+4L_2(id_2,\sigma_2^2)+L_2((i+1))d_2,\sigma_2^2\right] \qquad \ldots (15)$$

حيث أن:

$$v_j(\sigma_j^2)=(\sigma_j^2)^{-(\frac{v_j}{2}+1)}\exp-\frac{s_{Tj}^2}{2\sigma_j^2} \quad for \ \ j=1,2$$

بذلك يكون التوزيع التنبؤي المشترك للمشاهدتين المستقبليتين x_{n+1},z_{n+1} معرفاً على وفق المعادلة الآتية:

$$\ldots (16)$$

$$f(x_{n+1},z_{n+1}\mid X,Z)=\int_0^\infty\int_0^\infty f(x_{n+1},z_{n+1}\mid\theta_1,\theta_2,\sigma_1^2,\sigma_2^2)p(\sigma_1^2,\sigma_2^2\mid X,Z)d\sigma_1^2 d\sigma_2^2$$

يحسب العامل الأول من الطرف الأيمن في المعادلة (16) على وفق المعادلة الآتيــة: (سعيد (2005))[1]

$$f(x_{n+1}, \underline{z}_{n+1} | \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) = f(x_{n+1} | \underline{z}_{n+1}, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) p(\underline{z}_{n+1} | \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$$

$$= \prod_{j=1}^{2} (f_j(x_{n+1} | \theta_j, \sigma_j^2). \left(\frac{1}{F_j(b_j - a_j)} \right)^{z_{n+1,j}} P^{z_{n+1,j}} (1-P)^{1-z_{n+1,j}}$$

$$= \prod_{j=1}^{2} \left[\frac{P_j}{F_j(b_i - a_j)} . \frac{1}{\sqrt{2\pi}\sigma_j} \exp - (\frac{1}{2\sigma_j^2} .(x_{n+1} - \theta_j)^2)^{z_{n+1,j}} \right] \quad \ldots (17)$$

بعد تعويض المعادلتين (15) و (17) في المعادلة (16) نتوصل إلى التوزيع التنبؤي المشترك للمشاهدتين المستقبليتين x_{n+1}, z_{n+1} مرة أخرى تطبق قاعدة Simpson's على المعادلة الناتجة والتعويض بالقيم الممكنة للمشاهدة z_{n+1}، نحصل على التوزيع التنبؤي للمشاهدة المستقبلية x_{n+1} وتتكون من جزأين، كل جزء موزون بنسبة عائديه مشاهدات العينة إلى كل مركبة من مركبتي التوزيع الطبيعي المختلط بالمركبتين المبتورتين من طرفين والمعرفة في المعادلة (5).

5.الاستنتاجات

لقد توصلنا في هذا البحث إلى الاستنتاجين الآتيين:

1. إن استخدام البيانات الكاملة في أسلوب بيز سهل كثيراً عملية إيجاد المقدرات لمعلمات التوزيع الطبيعي ذي المركبات المبتورة.
2. إن مقدر بيز لمعلمة القياس θ_j يتكون من جزأين ،الأول يمثل مقدر بيز لتلك المعلمة في حالة عدم البتر، والثاني يمثل قيمة دالة المركبة عند نقطتي البتر.

6. التوصيات

1. نوصي باستخدام توزيعات سابقة مختلطة عن معلمات التوزيع الطبيعي المختلط ذي المركبات المبتورة وإيجاد المقدرات ومن ثم مقارنتها مع ما توصلنا إليه في هذا البحث.
2. استخدام إحدى طرائق معاينة جيبس في تحديد الثابت الطبيعي للتوزيع اللاحق المشترك لمعلمات التوزيعات المختلطة ذات المركبات المبتورة بديلاً عن استخدام الطرائق العددية في إيجاد ذلك الثابت.

7.المصادر

1. سعيد،هيفاء عبد الجواد (2005)،"تقدير معلمات التوزيعات المختلطة وتطبيقاتها على بيانات حديثي الولادة في محافظة نينوى"،أطروحة دكتوراه(غير منشورة)، كلية علوم الحاسبات والرياضيات،جامعة الموصل.

2. Cohen, A.C. (1960), "**Estimation in the Truncated Poisson when Zero and some Ones are Missing**", Journal of the American Statistical Association, Vol. 55, pp 342-348.

3. Cohen, A.C. (1966), "**A Note on Certain Discrete Mixed Distribution**", Biometrics, Vol. 22, pp 566-572.

4. Corduneanu, A. & Bishop, C.M. (2001), **"Variational Bayesian Model Selection for Mixture Distributions",** Artificial Intelligence and Statistics, T. Jaakkola and T. Richardson (Eds) pp 27-34, Morgan Kaufmann.

5. David, F. N. & Johnson, N.L. (1952), "The **Truncated Poisson**", Biometrics, Vol. 8, pp 275-285.

6. Love, C.E. & Rainville, E.D. (1969), "**Differential and Integral calculus**", New York, USA., The Macmillan Company.