مجلة البحوث المالية والتجارية
المجلد (22) – العدد الثاني – إبريل 2021

محاكاة توزيعات بواسون المختلطة المحدود لنمذجة عدد المطالبات التأمينية

Simulation of finite mixture Poisson distributions for modeling number of insurance claims

الباحث/ احمد رمضان محمد المتولى
المدرس المساعد بقسم الاحصاء والرياضيات والتأمين
كلية التجارة- جامعة بورسعيد

اشراف

أ.د/ محمد المهدى محمد على
استاذ الرياضيات والاحصاء الاكتوارى المتفرغ
كلية التجارة–جامعة بورسعيد

أ.م.د/ رانيا احمد حامد
استاذ الاحصاء التطبيقى المساعد
كلية التجارة–جامعة بورسعيد

رابط المجلة: https://jsst.journals.ekb.eg/

الملخص:

تهدف هذه الورقة البحثية الى تسليط الضوء حول احد الموضوعات الهامة فى مجال الاحصاء والمتمثلة فى توزيعات الخليط المحدودة باستخدام خوارزميتى تقدير ، هما : خوارزمية تعظيم التوقع Expectation Maximization (EM) Algorithm وخوارزمية تقدير بايز المحسن التقريبى The Rough-Enhanced-Bayes Mixture (REBMIX) algorithm من خلال بناء توزيع بواسون المختلط بمكونين وبثلاث بمكونات وبأربع مكونات ومن ثم مقارنة النتائج الخاصة بالتقدير والمستخرجة من كلتا الخوارزميتين عن طريق المحاكاه باستخدام كل من معيار المعلومات البيزى Bayesian information criterion (BIC) وقيمة لوغاريتم الامكان Log likelihood (LL). وقد توصلت الدراسة الى أنه وبالنظر الى معايير المقارنة فإن التقديرات المستخرجة باستخدام خوارزمية REBMIX تعطى نتائج أفضل فى التقدير مع كل الحالات المختلفة للمحاكاة.

الكلمات المفتاحية : توزيع بواسون ، التوزيعات المختلطة المحدودة ، تقديرات الامكان الأعظم ، خوارزمية تعظيم التوقع ، خوارزمية تقدير بايز المحسن التقريبى ، المحاكاة.

Abstract:

This paper aims to shed light on one of the most important topic in statistics, this topic is could finite mixture distributions. By using two estimation algorithms, namely: The Expectation Maximization (EM) Algorithm and The Rough-Enhanced-Bayes Mixture (REBMIX) algorithm. By building a mixture distribution of Poisson with two, three, and four components. Then, the results of the estimation and extracted from both algorithms are compared by simulation using: Bayesian information criterion (BIC) and the value of the log-likelihood (LL). The study found that, The estimates extracted using the REBMIX algorithm give better results in the estimation with all the different cases of simulation.

Keywords: Poisson Distribution, Finite Mixture Distributions, Maximum Likelihood Estimation, Expectation Maximization Algorithm, The Rough-Enhanced-Bayes Mixture Algorithm, Simulation.

1. المقدمة Introduction:

ان الزيادة الكبيرة فى حجم البيانات والتفاوت الشديد فى القيم الممثلة لها ، يدفعنا الى الحاجة الى استخدام طريقة معقولة لوصف عدم التجانس غير المشاهد فى المجتمع. وتعتبر نماذج التوزيعات المختلطة اسلوب احصائى قوى يوفر أساساً لوصف خصائص تلك البيانات. وتستخدم التوزيعات المختلطة المحدودة اذا كان المجتمع الأصلى مُعَروف بدقة والهدف هو نمذجة مجتمعات فرعية محدودة بشكل جيد ، لكن تكلفة الحصول على كل منها بصورة فرعية كبيرة للغاية. وحيث أن بيانات المطالبات التأمينية تتميز بالتفاوت الشديد فى قيمها ، وعليه فان التقدير الجيد ونمذجة المطالبات يؤدى الى تسعير جيد للتأمين على الممتلكات.

ويتمثل الحافز الرئيسى للقيام بهذه الورقة البحثية فى محاولة البحث عن طريقة جديدة لتقدير نموذج للتوزيع المختلط لبيانات عدد المطالبات التأمينية يعتمد على تقسيم البيانات الى مجموعات متجانسة ومن ثم دراسة الخصائص الفردية الممثلة لهذه البيانات اعتماداً على التوزيع التقليدى الممثل لكل مجموعة. ان الاجراء الأكثر شيوعا للوصول الى الأمثلية عند تقدير النماذج المختلطة المحدودة هو استخدام خوارزمية Expectation Maximizatio (EM)algorithm وهو اجراء يهدف الى الحصول على اقصى قيمة للوغاريتم الامكان المتوقع. الا أن الهدف الرئيسي لهذه الورقة البحثية يتمثل فى محاولة بناء واقتراح نماذج جديدة للمطالبات التأمينية تستند الى توزيعات خليط محدودة لمجموعة بيانات عدد المطالبات الفعلية. وتسعى الدراسة من خلال استخدام التوزيعات المختلطة المحدودة الى اثبات صحة أو خطأ الفرض القائل: " إن النماذج المبنية باستخدام التوزيعات المختلطة والمقدرة باستخدام طريقة REBMIX_algorithm تعطى نتائج أفضل من النماذج المبنية باستخدام التوزيعات المختلطة اعتماداً على تقديرات طريقة EM_algorithm عند التقدير والتنبؤ بعدد المطالبات المتوقع".

2. توزيع بواسون The Poisson Distribution

يعتبر توزيع بواسون احد أشهر التوزيعات المتقطعة والتى تستخدم لتمثيل عدد المطالبات التأمينية ، ويصلح هذا التوزيع للحوادث النادرة الوقوع أي عندما يكون عدد المشاهدات (n) كبيراً جدا بينما يكون احتمال تحقق الحادث صغيرا جدا. وتكون دالة pdf لهذا التوزيع على الصورة: (Klugman & et al, 2012) $f\left(x\right)=e^{-\lambda}\lambda^{x}/x!\ ,\ \ x=0,1,2,\ldots,\infty$ حيث تمثل λ مقدار ثابت يمثل متوسط عدد الحوادث خلال فترة زمنية محددة $\lambda>0$ و e مقدار ثابت يمثل لوغاريتم الاساس الطبيعى.

3. تقدير نماذج الخليط المحدود Estimation in Finite Mixture Models

لنماذج الخليط المحدود طرقا متعددة للتقدير منها ما يعتمد على العمليات التكرارية ومنها ما يستخدم الأسلوب البيزى أو العزوم أو من خلال تطبيق الأدوات الرسومية مثل طريقة المسافة الصغرى (Grün & Leisch, 2007). فى هذه الورقة البحثية سوف نتكلم بايجاز عن التقدير اعتمادا على طريقة التقدير باستخدام تقديرات الإمكان الأعظم Maximum Likelihood Estimations (MLE) وخوارزمية أقصى توقع Expectation-Maximization (EM) Algorithm مع طرح لفكرة استخدام خوارزمية جديدة تتلاشى بعض من عيوب تقديرات (EM) Algorithm هذه الخوارزمية هى (REBMIX) Algorithm.

4.تقدير الامكان الأعظم Maximum Likelihood Estimation

بافتراض ان قيم العينة $\{x_1, x_2, ..., x_n\}$ تمثل تغيرات عشوائية مستقلة للبيانات الخاصة بتوزيع العينة. بمعنى آخر، بيانات العينة لها توزيع مشترك $f_x(x_1, x_2, ..., x_n|\theta)$ مع المعلمة θ وبافتراض ان هذه التغيرات مستقلة وموزعة بصورة متماثلة independent and identically distributed (iid) (Ghojogh & et al, 2019) ، وباستخدام قاعدة بايز ، حيث:

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

حيث $P(A)$ يمثل الاحتمال القبلى ، $P(B)$ يمثل الاحتمال الهامشى ، بينما $P(B|A)$ يمثل الاحتمال البعدى و $P(A,B)$ يمثل الاحتمال المشترك لكل من A و B. مما سبق ، نجد أن:

$$L(\theta|x_1, x_2, ..., x_n) = \frac{f_x(x_1, x_2, ..., x_n|\theta)\pi(\theta)}{f_x(x_1, x_2, ..., x_n)}$$

وحيث أن الهدف من تطبيق تقدير الأمكان الأعظم MLE هو ايجاد المعلمة θ والتى لديها أكبر احتمال ممكن ، حيث:

$$\hat{\theta} = \arg \max_{\theta} L(\theta)$$

من التعريف السابق يمكن كتابة دالة الامكان على الصورة:

$$L(\theta|x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n; \theta) \overset{(a)}{=} \prod_{i=1}^{n} f(x_i; \theta)$$

حيث يتم وضع (a) لأن القيم $\{x_1, x_2, ..., x_n\}$ موزعة كما سبق (iid). علما بأن بعض الأدبيات الأخرى تختصر الرمز السابق للتعبير عن دالة الامكان على الصورة $L(\theta)$ وذلك للتبسيط.(Filho, 2008) ايضا ، سوف يستخدم الرمز $LL(\theta)$ للتعبير عن لوغاريتم دالة الامكان وعليه ، تكون قيمة لوغاريتم دالة الامكان على الصورة (Yen, Ismail, & Hamzah, 2014):

$$LL(\theta) = \log L(\theta) = \log \prod_{i=1}^{n} f(x_i;\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

وللحصول على اقصى قيمة لدالة الامكان وبالتالى الحصول على افضل تقديرات للنموذج ، نقوم بمساواة المشتقة الأولى الجزئية لدالة الامكان بالنسبة للمعلمات بالصفر ، أى:

$$\frac{\partial l}{\partial \theta_i} = 0$$

وبحل المعادلات المتكونة يمكننا من خلالها تقدير قيم المعلمات. وعند استخدامنا للتوزيعات المختلطة ، تكون الأمور اكثر تعقيدا ، حيث ان المعادلات المتكونه قد لا تكون قابلة للحل بصورة صريحة للحصول على المعلمات θ بل يلزم الاعتماد على تقنيات تستخدم عمليات تكرارية من اجل الحصول على قيمة عظمى لدالة الامكان.

4.1 خوارزمية تعظيم التوقع EM Algorithm

نشأت خوارزمية تعظيم التوقع على يد Dempster, Laird and Rubin (1977) (Dempster, Laird, & Rubin, 1977)، اليوم وبعد اكثر من اربعين عاما لا تزال واحدة من اهم طرق التعرف على الانماط الاحصائية (Pani´c & et al, 2020). هذه الورقة البحثية تحاول تقديم شرح مختصر عن تلك الخوارزمية (Peter & Jones, 2003). وتمثل خوارزمية تعظيم التوقع (EM) طريقة عامة لإيجاد مقدرات الإمكان الأعظم فى حال وجود قيم مفقودة Missing Values أو متغيرات كامنة Latent Variable (Filho, 2008). ففى بعض الاحيان لا يمكن مشاهدة البيانات بصورة كاملة ، بمعنى آخر، البيانات تكون مكتملة. ومثال على ذلك بافتراض جمع بيانات حول مرض معين ، حيث لم يتم تسجيل بيانات حول شدة المرض وانما تسجيل وجود أو عدم وجود المرض. اى أن التعبير هنا عن البيانات اما بصفر والتى تمثل عدم توافر الظاهرة أو قيمة أكبر من الصفر فى حال توافر الظاهرة. من هنا نجد أن البيانات لا تعطينا صورة كاملة حول الظاهرة حيث انه عندما تكون قيمة $x > 0$ فإننا لا نعرف بالتحديد هل تساوى 5 أم تساوى 100. فى تلك الحالة لا يمكننا تطبيق MLE بشكل مباشر لأننا فى هذه الحالة لا نملك معلومات كاملة حول الظاهرة وبالتالى توجد بعض القيم المفقودة Missing Values. وتنقسم خوارزمية EM الى خطوتين رئيسيتين (Chris & Raftery, 2017). **الخطوة الأولى E-Step:** وتهدف الى الحصول على القيمة المتوقعة لمعادلة الامكان حيث يتم اخذ التوقع الخاص بلوغاريتم الامكان. من أجل الحصول على تقديرات متوسطة للقيم المفقودة $D^{(Miss)}$. وبافتراض أن $Q(\theta)$ ترمز الى توقع الامكان وذلك فيما يتعلق بقيم $D^{(Miss)}$ حيث:

$$Q(\theta) = E_{D^{(Miss)}|D^{(obs)},\theta}\left[LL(\theta)\right]$$

من التوقع السابق يتبين أن هذا التوقع مشروط بكل من $\theta, D^{(obs)}$ لذا يتم التعامل معها هنا كثوابت وليست متغيرات. اما **الخطوة الثانية M_Step:** تعتمد على استخدام منهجية MLE ، حيث يتم استبدال لوغاريتم الامكان بالتوقع ، بمعنى أخر استبدال المقدار $\max_{\theta} L(\theta)$ فى المعادلة:

$$\hat{\theta} = \arg \max_{\theta} Q(\theta)$$

وباعادة انتاج الخطوتين السابقتين بشكل متكرر يتم الوصول الى تقارب حول المعلمات المقدرة $\hat{\theta}$.

5. تقدير معلمات التوزيعات المختلطة

ان الهدف من تكوين توزيع خليط مناسب هو الحصول على قيم لمعلمات هذا التوزيع. لتوضيح ذلك رياضيا ، وبالنظر الى توزيعات الخليط لمكونين وبافتراض ان دالتى كثافة الاحتمال للمكونين هما $g_1(x;\theta_1)$ و $g_2(x;\theta_2)$ مع ملاحظة ان كلا التوزيعين ليسا بالضرورة ينتميان الى نفس العائلة. وعليه نجد ان دالة الكثافة الاحتمالية لخليط مكون من توزيعين هو (Fraley & Raftery, 2002):

$$f(x;\theta) = \sum_{i=1}^{2} w_i g_i(x;\theta_1,\theta_2) = w g_1(x;\theta_1) + (1-w) g_2(x;\theta_2)$$

حيث تمثل $(w, 1-w)$ نسبة الخليط لكل مكون من المكونين السابقين على الترتيب حيث تعبر w عن معلمة الخليط Mixing Probability. علما بان مجموع النسبتين يجب ان يساوى الواحد الصحيح. والان سوف يتم شرح كيفية ايجاد التوزيعات المختلطة لتوزيع مكون من خليط من توزيعين. ويمكن كتابة كل من دالة الامكان للخليط على الصورة (Frempong & et al., 2017):

$$L(\theta_1,\theta_2) = f(x_1,x_2,...,x_n;\theta_1,\theta_2) \overset{(a)}{=} \prod_{i=1}^{n} f(x_i;\theta_1,\theta_2) = \prod_{i=1}^{n} w g_1(x_i;\theta_1) + (1-w) g_2(x_i;\theta_2)$$

مع ملاحظة انه تم وضع (a) لأن القيم $\{x_1,x_2,...,x_n\}$ موزعة بصورة مستقلة ومتماثلة (iid).

ومنها ، يصبح لوغاريتم دالة الامكان على الصورة:

$$LL(\theta_1,\theta_2) = \sum_{i=1}^{n} \log\left[w g_1(x_i;\theta_1) + (1-w) g_2(x_i;\theta_2)\right]$$

بسبب وجود عمليات جمع داخل اللوغاريتم ، فمن الصعب ايجاد الامثلية للوغاريتم دالة الامكان فى المعادلة (Fraley & Raftery, 2002)، لذا يتم عمل التقسيم التالى:

$$\Delta_i = \begin{cases} 1 & if \;\; x_i \;\; belongs \; to \;\; g_1(x_i;\theta_1) \\ 0 & if \;\; x_i \;\; belongs \; to \;\; g_2(x_i;\theta_2) \end{cases}$$

مما سبق يمكن الحصول على احتمال ان تنتمى المشاهدة الى احدى التوزيعين ، وفقا لما يلى:

$$\begin{cases} P(\Delta_i = 1) = w \\ P(\Delta_i = 0) = 1-w \end{cases}$$

ومن ثم فإن لوغاريتم الامكان يمكن كتابته على الصورة:

$$LL(\theta_1,\theta_2)=\begin{cases}\sum_{i=1}^{n}\log[wg_1(x_i;\theta_1)] & if \quad \Delta_i=1\\ \sum_{i=1}^{n}\log\left[(1-w)g_2(x_i;\theta_2)\right] & if \quad \Delta_i=0\end{cases}$$

ويمكن اعادة كتابة المعادلة السابقة على النحو التالى:

$$LL(\theta_1,\theta_2)=\sum_{i=1}^{n}\left[\Delta_i\log[wg_1(x_i;\theta_1)]+(1-\Delta_i)\log\left[(1-w)g_2(x_i;\theta_2)\right]\right]$$

حيث يعبر الرمز Δ_i عن المعلومات غير الكاملة (المفقودة) والتى تأخذ القيمتين 1 أو 0 لكل مشاهدة من المشاهدات x_i. وبالتالى فإن استخدام خوارزمية EM تمكننا من محاولة تقدير المعلمات اعتمادا على التوقع ، وعندها نجد انه فى الخطوة الأولى فى خوارزمية EM والخاصة بـ E-Step يمكن كتابتها على الصورة:

$$Q(\theta_1,\theta_2)=\sum_{i=1}^{n}\left[E\left[\Delta_i|X,\theta_1,\theta_2\right]\log[wg_1(x_i;\theta_1)]+E\left[1-(\Delta_i)|X,\theta_1,\theta_2\right]\log\left[(1-w)g_2(x_i;\theta_2)\right]\right]$$

وحيث أن التغيرات السابقة خطية فيما يتعلق بـ Δ_i ، لذا تم اخذ اللوغاريتمين. ولحساب الخطوة الأولى لخوارزمية تعظيم التوقع E-Step بفرض أن $\hat{\gamma}_i=E\left[\Delta_i|X,\theta_1,\theta_2\right]$ ، حيث $\hat{\gamma}_i$ تسمى دالة Responsibility لـ x_i او ما يطق عليها الوزن النسبى لكل توزيع (Fraley & Raftery, 2002). وبما أن Δ_i يأخذ احدى قيمتين {0,1} فإن:

$$\begin{aligned}E\left[\Delta_i|X,\theta_1,\theta_2\right]\log[wg_1(x_i;\theta_1)]&=0\times P(\Delta_i=0|X,\theta_1,\theta_2)+1\times P(\Delta_i=1|X,\theta_1,\theta_2)\\&=P(\Delta_i=1|X,\theta_1,\theta_2)\end{aligned}$$

وبالتعويض يكون لدينا الشكل التالى:

$$P(\Delta_i=1|X,\theta_1,\theta_2)=\frac{P(X,\theta_1,\theta_2,\Delta_i=1)}{P(X;\theta_1,\theta_2)}=\frac{P(X,\theta_1,\theta_2|\Delta_i=1)P(\Delta_i=1)}{\sum_{j=0}^{1}P(X,\theta_1,\theta_2|\Delta_i=1)P(\Delta_i=j)}$$

من هنا يمكن ايجاد دالة Responsibility أو ما تسمى دالة الاحتمال البعدى (Ng & et al., 2019) ، على الصورة:

$$\hat{\gamma}_i=\frac{\hat{w}g_1(x_i;\theta_1)}{\hat{w}g_1(x_i;\theta_1)+(1-\hat{w})g_2(x_i;\theta_2)}$$

ويكون التوقع على الصورة:

$$Q(\theta_1,\theta_2)=\sum_{i=1}^{n}\left[\hat{\gamma}_i\log w+\hat{\gamma}_i\log g_1(x_i;\theta_1)+(1-\hat{\gamma}_i)\log(1-w)+(1-\hat{\gamma}_i)\log g_2(x_i;\theta_2)\right]$$

وبتطبيق الخطوة الثانية M-Step ، نجد أن:

$$\hat{\theta}_1,\hat{\theta}_2,\hat{w}=\arg\max_{\theta_1,\theta_2,w}Q(\theta_1,\theta_2,w)$$

وبايجاد المشتقة الأولى لدالة التوقع ومساوتها بالصفر ، نلاحظ ما يلى:

$$\frac{\partial Q}{\partial \theta_1}=\sum_{i=1}^{n}\left[\frac{\hat{\gamma}_i}{g_1(x_i;\theta_1)}\frac{\partial g_1(x_i;\theta_1)}{\partial \theta_1}\right]\overset{set}{=}0$$

$$\frac{\partial Q}{\partial \theta_2}=\sum_{i=1}^{n}\left[\frac{\hat{\gamma}_i}{g_2(x_i;\theta_2)}\frac{\partial g_2(x_i;\theta_2)}{\partial \theta_2}\right]\overset{set}{=}0$$

$$\frac{\partial Q}{\partial w}=\sum_{i=1}^{n}\left[\hat{\gamma}_i\left(\frac{1}{w}\right)+\left(1-\hat{\gamma}_i\right)\left(\frac{-1}{1-w}\right)\right]\overset{set}{=}0\Rightarrow \hat{w}=\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i$$

وبحل المعادلات الثلاث السابقةبصورة تكرارية وصولا الى التقارب يتم الحصول على تقديرات لكل من $\left(\hat{\theta}_1,\hat{\theta}_2,\hat{w}\right)$.

4.1 خليط من توزيعى بواسون Mixture of Two Poisson

بافتراض ان لدينا بيانات تمثل عدد مرات تكرار حادث خلال فترة سريان وثيقة التأمين وأنه يمكن التعبير عن عدد مرات تكرار الحادث باستخدام توزيع مختلط مكون من توزيعى بواسون ، وأن دالتى كثافة احتمال لكلا التوزيعين بمعلمتين مختلفتين حيث :

$$g_1\left(x_i;\lambda_1\right)=\frac{\lambda_1^{x_i}e^{-\lambda_1}}{x_i\,!}\qquad ,\quad g_2\left(x;\lambda_2\right)=\frac{\lambda_2^{x_i}e^{-\lambda_2}}{x_i\,!}$$

وباستخدام المعادلة السابقة يمكن اعادة كتابة دالة كثافة الاحتمال للتوزيع المختلط على الصورة:

$$f\left(x_i;\lambda_1,\lambda_2\right)=w\,\frac{\lambda_1^{x_i}e^{-\lambda_1}}{x_i\,!}+\left(1-w\right)\frac{\lambda_2^{x_i}e^{-\lambda_2}}{x_i\,!}$$

وعليه فإن:

$$\hat{\gamma}_i=\frac{\hat{w}\,\dfrac{\lambda_1^{x_i}e^{-\lambda_1}}{x_i\,!}}{\hat{w}\,\dfrac{\lambda_1^{x_i}e^{-\lambda_1}}{x_i\,!}+\left(1-\hat{w}\right)\dfrac{\lambda_2^{x_i}e^{-\lambda_2}}{x_i\,!}}$$

من هنا يمكن ايجاد دالة التوقع للتوزيع:

$$Q\left(\lambda_1,\lambda_2\right)=\sum_{i=1}^{n}\left[\hat{\gamma}_i\log w+\hat{\gamma}_i\log\left(\frac{\lambda_1^{x}e^{-\lambda_1}}{x\,!}\right)+\left(1-\hat{\gamma}_i\right)\log\left(1-w\right)+\left(1-\hat{\gamma}_i\right)\log\left(\frac{\lambda_2^{x}e^{-\lambda_2}}{x\,!}\right)\right]$$

$$=\sum_{i=1}^{n}\left[\begin{array}{l}\hat{\gamma}_i\log w+\hat{\gamma}_i\left(x_i\log(\lambda_1)-\lambda_1-\log(x_i)\right)+\left(1-\hat{\gamma}_i\right)\log\left(1-w\right)\\+\left(1-\hat{\gamma}_i\right)\left(x_i\log(\lambda_2)-\lambda_2-\log(x_i)\right)\end{array}\right]$$

وباستخدام المشتقة الأولى للتوقع ومساواتها بالصفر يمكن الحصول على تقديرات للمعلمات:

$$\frac{\partial Q}{\partial \lambda_1}=\sum_{i=1}^{n}\left[\hat{\gamma}_i\left(\frac{x_i}{\lambda_1}-1\right)\right]=0\qquad\Rightarrow\qquad \lambda_1=\frac{\sum_{i=1}^{n}\hat{\gamma}_i x_i}{\sum_{i=1}^{n}\hat{\gamma}_i}$$

وبالمثل نجد ان:

$$\frac{\partial Q}{\partial \lambda_2}=\sum_{i=1}^{n}\left[\left(1-\hat{\gamma}_i\right)\left(\frac{x_i}{\lambda_2}-1\right)\right]=0\quad\Rightarrow\quad \lambda_2=\frac{\sum_{i=1}^{n}\left(1-\hat{\gamma}_i\right)x_i}{\sum_{i=1}^{n}\left(1-\hat{\gamma}_i\right)}$$

وبحل المعادلات السابقة بصورة تكرارية باستخدام خطوات بناء خوارزمية EM لمكونين وبتالى الحصول على تقديرات لـ $\left[w,\lambda_1,\lambda_2\right]$.

4.2 خليط من عدة توزيعات لبواسون Mixture of Several Poisson

باستخدام دالة كثافة الاحتمال لبواسون يمكن كتابة كثافة الخليط (Ghojogh & et al, 2019) ، على الصورة:

$$f\left(x;\lambda_1,...,\lambda_K\right)=\sum_{k=1}^{K} w_k \frac{e^{-\lambda_k} \cdot \lambda_k^{X}}{X!}$$

وبالتالى فإن الاحتمال البعدى يكون على الصورة:

$$\widehat{\gamma}_{i,k}=\frac{w_k \dfrac{e^{-\widehat{\lambda}_k} \cdot \widehat{\lambda}_k^{X_i}}{X_i!}}{\sum_{k'=1}^{K} \widehat{w}_{k'} \dfrac{e^{-\widehat{\lambda}_{k'}} \cdot \widehat{\lambda}_{k'}^{X_i}}{X_i!}}$$

وبالتالى نجد أن دالة التوقع تكون على الصورة:

$$Q\left(\lambda_1,...,\lambda_k\right)=\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\widehat{\gamma}_{i,k}\log\left(w_k\right)+\widehat{\gamma}_{i,k}\left(-\lambda_k+X_i\log\left(\lambda_k\right)-\log\left(X_i!\right)\right)\right]$$

وباضافة المقدار Lagrangian والذى يعتمد على معامل لاجرانج نجد أن:

$$LL\left(\lambda_1,...,\lambda_K,w_1,...,w_k,\alpha\right)=\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\widehat{\gamma}_{i,k}\log w_k+\widehat{\gamma}_{i,k}\left(-\lambda_k+X_i\log\left(\lambda_k\right)-\log\left(X_i!\right)\right)\right]-\alpha\left(\sum_{k=1}^{K}w_k-1\right)$$

و يمكن تقدير المعلمات عن طريق ايجاد المشتقة الأولى للوغاريتم دالة الامكان على الصورة:

$$\frac{\partial LL\left(\lambda_1,...,\lambda_K,w_1,...,w_k,\alpha\right)}{\partial\lambda_k}=\sum_{i=1}^{n}\left[\widehat{\gamma}_{i,k}\left(\frac{X_i}{\lambda_k}-1\right)\right]\overset{set}{=}0$$

$$\widehat{\lambda}_k=\frac{\sum_{i=1}^{n}\widehat{\gamma}_{i,k}x_i}{\sum_{i=1}^{n}\widehat{\gamma}_{i,k}}$$

وبحل المعادلة السابقة بصورة تكرارية مع معادلة الأوزان يمكن الحصول على تقديرات للمعلمات $\left(\lambda_1,...,\lambda_K,w_1,...,w_k\right)$.

6. خوارزمية REBMIX:

ظهرت هذه الخوارزمية على يد كل من Nagode and Fajdiga (1998) بهدف تجنب بعض العيوب التى ظهرت مع استخدام طريقة EM (Nagode, 2013). حيث يتطلب منهج الخليط المحدود الكلاسيكى استخدام خوارزمية EM فى حل معادلة الامكان ، وعليه يمكن تقدير جميع المعلمات غير المعروفة فى نموذج الخليط بصورة تكرارية وبالتالى تضمن خوارزمية EM ان قيمة الامكان تزداد ولكن بشكل رتيب The Likelihood Increase Monotonically أى على وتيرة واحدة ولا يتغير تقريبا (Nagode & Fajdiga, 2011). أيضا، فبدلا من تحديد عدد المكونات والتخمينات الأولية لأوزان المكونات والمعلمات ، تضمن معلمة الإدخال Input parameter لخوارزمية REBMIX – والتى تعتبر ايضا بمثابة معلمات تمهيد Smoothing Parameter – اجراء تقديرات للكثافات التجريبية بصورة اكثر دقة من خوارزمية EM. من هنا ،

يمكن تلخيص عيوب استخدام خوارزمية EM فى النقاط التالية (Nagode , 2013) & Panić) at el., 2020):

أ. تتقارب خوارزمية EM الى أقصى قيمة محلية لدالة الامكان بصورة كبيرة.

ب. غالبا ما يكون هناك العديد من الحلول المثلى المحتملة بالقرب من الحلول التى تم الحصول عليها باستخدام التخمين الأولى للحل.

ج. فى الغالب تفترض معايير اختيار النموذج انه يمكن الحصول على حل أمثل عام للوغاريتم دالة الامكان ، الا أن تحقيق ذلك يعد امراً صعباً من الناحية الحسابية.

د. لا تحتوى بعض مناطق البحث على حلول واعدة محتملة ومن ثم يتم اهدار مزيد من الوقت فى البحث عن الحل الأمثل فى المناطق غير الواعدة حسابياً.

من هنا يمكن تلخيص هذه العيوب فى أن مجموعة المعلمات الكاملة غير المعرفة يتم تقديرها فى وقت واحد فى كل خطوه من خطوات التكرار عند استخدام خوارزمية EM ، ومن ثم فإنه يترتب على ذلك احتمال حدوث تعظيم محلى زائف اذا زاد عدد المكونات أو عدد الأبعاد. ويحسب احتمال التعظيم المحلى الزائف من خلال قسمة عدد التقديرات بدون تعظيم محلى زائف على العدد الاجمالى للتقديرات (Nagode & Fajdiga, 2011).

وتعتبر خوارزمية REBMIX اجراء عددى تكرارى يعتمد على مجموعة من الخطوات ، هى:

1- تحديد كثافات تجريبية Empirical densities لمجموعة البيانات المشاهدة ، ولعل من اشهر تلك الطرق استخدام مدخل الرسم البيانى وخاصة المدرج التكرارى أو نافذة Parzen window أو منهج الجار الأقرب K-nearest neighbor approach.

2- بناء على هذه الكثافات التجريبية يمكن تصور شكل النمط العام Global Mode.

3- بمجرد معرفة النمط العام وكثافته التجريبية ، يمكن تقدير معلمات المكون التقريبية The Rough component parameters لكثافات المكون التنبؤية.

4- يتم تجميع المشاهدات فى مجموعتين فرعيتين بصورة تدريجية ، المجموعة الأولى مرتبطة بكثافة المكون التنبؤية والأخرى تمثل البواقى ، وهى المشاهدات التى تنتمى الى بقية المكونات. ومن ثم يتم تقدير وزن الخليط لكثافة المكون التنبؤية.

5- يتم تكرار الخطوتين الثالثة والرابعة حتى يتم استيفاء معيار التقارب The convergence criterion ، ثم يتم الوصول الى تقدير محسن لمعلمة المكون وكذلك اوزان المكونات اعتمادا على طريقة MLE.

6- يتم تكرار الخطوات من الثانية وحتى الخامسة بشكل تكرارى حتى يتم تحسين معيار المعلومات المعلومات Information Criterion وبذلك يتم تحديد عدد المكونات ومعلمات الخليط وأوزان الخلط.

7- اخيرا، يتم توزيع المشاهدات المتبقية والتى تمثل البواقى بين المكونات الموجودة بواسطة قاعدة بايز Bayes rule ، ومن ثم ضبط معلمات الخليط المحدود بدقة.

من العرض السابق ، يظهر أحد أهم الفروق الواضحة بين خوارزمية EM وخوارزمية REBMIX وهى أن معلمات المكون وأوزان الخليط وعدد المكونات تتحدد فى خوارزمية REBMIX على التوالى وليست بصورة آنية مثل خوارزمية EM.

ومن ثم فإن احتمال حدوث تعظيم محلى زائف يعتبر غير موجود اذا زاد عدد المكونات (Nagode & Fajdiga, 2011).

7. دراسة المحاكاة:

اعتمدت هذه الورقة البحثية على استخدام حزم برنامج R باصدار R i386 4.0.3 فى عملية المحاكاة. ان عملية المحاكاة هذه تهدف الى التحقق من صحة تقديرات المعلمات لتوزيعات بواسون المختلطة محل الدراسة. فى هذه الدراسة تم استخدام كل من لوغاريتم الامكان السالب ومعيار المعلومات البيزى كمعيارى مقارنة رئيسيين حيث كلما كانت قيمة لوغاريتم الامكان المطلقة اكبر ما يمكن دل ذلك على كفاءة المقدرات ، أيضا فإن معيار المعلومات البيزى Bayesian information criterion (BIC) والذى يطلق عليه احيانا معيار شوارتز Schwarz criterion (also SBC, SBIC) من المعايير المطبقة فى اختيار النموذج الأفضل من بين مجموعة من النماذج ، حيث يعتمد جزئيا على دالة الامكان ، كما أنه وثيق الصلة بمعيار المعلومات لأكايكى Akaike information criterion (AIC) ، أيضا فقد تم تفضيل استخدام معيار المعلومات البيزى عن معيار المعلومات لأكايكى لأن فترة The penalty term او ما تسمى بفترة الجزاء أكبر فى معيار المعلومات البيزى عنها فى معيار أكايكى وبالتالى فجودة نتائجه أفضل (2021 ,.......). ايضا، فالصيغة الحسابية لمعيار المعلومات البيزى تعتمد على حساب الجزاء أو العقوبة باستخدام لوغاريتم الامكان ، حيث تأخذ العقوبة الشكل التالى:

$$BIC(M) = 2\log - likelihood_{\max}(M) - \log(n)\dim(M)$$

حيث يعبر M عن النموذج ، و dim(M) عن عدد المعلمات المقدرة فى النموذج ، و n تمثل حجم العينة ، ويتم اختيار النموذج الذى يعطى أكبر قيمة لـ BIC. كذلك ، فإن قيمة BIC تكون أفضل من AIC طالما أن حجم العينة $n \geq 8$ (Claeskens & Hjort, 2008).

8.تصميم المحاكاه:

تتم عملية المحاكاة باستخدام بيانات افتراضية حول أعداد المطالبات المتوقعه لفرع من فروع التأمين على الممتلكات من أحجام مختلفة من العينات وهى 500 و 1500 و 2500 مفردة بنسب تمثيل مختلفة. ويتم تكوين توزيعات مختلطة محدودة متماثلة من عدد من المكونات k مفترضة للدراسة ، حيث k=2,3,4. ويمكن عرض توزيعات الخليط المحاكاه وفقا للجدول التالى:

جدول (1): توزيعات الخليط المولدة لعملية المحاكاه بالنسبة لتوزيعات بواسون لتكرار الخسارة

عدد المكونات	توزيعات الخليط	المعلمات
2	Poisson	w, λ_1, λ_2
3	Poisson	$w_1, w_2, \lambda_1, \lambda_2, \lambda_3$
4	Poisson	$w_1, w_2, w_3, \lambda_1, \lambda_2, \lambda_3, \lambda_4$

أيضا تتم عملية المحاكاه مع افتراض أن اوزان للخليط على ثلاث حالات للأوزان ، هي:

جدول (2): اوزان الخليط الخاصة بنسب المكونات

متفاوته	متساوية	متقاربة
$w_1 = 0.2, w_2 = 0.8$	$w_1 = w_2 = 0.5$	$w_1 = 0.4, w_2 = 0.6$
$w_1 = 0.1, w_2 = 0.3, w_3 = 0.6$	$w_1 = w_2 = w_3 = 0.333$	$w_1 = 0.25, w_2 = 0.3, w_3 = 0.45$
$w_1 = 0.1, w_2 = 0.3, w_3 = 0.5, w_4 = 0.1$	$w_1 = w_2 = w_3 = w_4 = 0.25$	$w_1 = 0.2, w_2 = 0.3, w_3 = 0.3, w_4 = 0.2$

كذلك ، تم افتراض قيماً تقريبية للمعلمات الحقيقية من خلال النظر الى طبيعة بيانات المطالبة التأمينية ، حيث كانت النسب لتوزيعات الخليط وفقا للجدول التالي:

جدول (3): القيم الفعلية للمعلمات لتوزيعات الخليط المولدة لعملية المحاكاه

توزيعات الخليط	معلمات المكونات
Poisson(2)	$\lambda_1 = 2, \lambda_2 = 0.5$
Poisson(3)	$\lambda_1 = 2.5, \lambda_2 = 1.5, \lambda_3 = 0.5$
Poisson(4)	$\lambda_1 = 3.5, \lambda_2 = 2.5, \lambda_3 = 1, \lambda_4 = 0.5$

9.نتائج المحاكاه لتوزيعات بواسون المختلطة:

تمت عملية المحاكاه باستخدام توزيع بواسون ، ومن ثم حصلنا على تقديرات لمعلمات النموذج باستخدام كل من خوارزميتى EM و REBMIX حيث تم فرض قيد على عدد التكرارات بحث يكون عدد التكرارات مساوى لحجم العينة كما تم التحقق من التقارب بوضع قيد على عملية التقارب بحيث يتوقف الاختبار اذا كانت الفروق بين التقديرات المتتالية اقل من 10^{-4} ، كما أن الطريقة المتبعة في مرحلة ما قبل المعالجة بالنسبة لخوارزمية REBMIX تمثلت فى استخدام مدخل histogram ، ويمكن عرض النتائج التى تم التوصل اليها الدراسة ، بالجداول التالية:

جدول (4): نتائج المحاكاه لتقدير خليط من توزيع بواسون بحجم عينة (500) مفردة

الخليط	k	خوارزمية التقدير	المعلمات الفعلية	التقديرات (اوزان متفاوته)	LL	BIC	التقديرات (اوزان متساوية)	LL	BIC	التقديرات (اوزان متقاربة)	LL	BIC
Poisson	2			$w_1=0.2, w_2=0.8$			$w_1=w_2=0.5$			$w_1=0.4, w_2=0.6$		
		EM	w_1	0.2864	-621.3739	1261.392	0.6254	-766.4718	1551.587	0.5412	-727.4011	1473.446
			$\lambda_1=2$	1.7794			1.8652			1.7903		
			$\lambda_2=0.5$	0.4293			0.3032			0.3642		
		REBMIX	w_1	0.3912	-621.6628	1261.969	0.619	-766.4865	1551.617	0.997	-745.7316	1510.107
			$\lambda_1=2$	1.563			1.8686			1.1175		
			$\lambda_2=0.5$	0.3359			0.3241			72036		
	3			$w_1=0.1, w_2=0.3, w_3=0.6$			$w_1=w_2=w_3=0.333$			$w_1=0.25, w_2=0.35, w_3=0.4$		
		EM	w_1	0.2535	-676.4019	1383.877	0.4262	-805.2017	1641.477	0.3589	-773.7113	1578.496
			w_2	0.249			0.3994			0.3734		
			$\lambda_1=2.5$	1.803			1.8059			1.9904		
			$\lambda_2=1.5$	1.0848			1.7809			1.4452		
			$\lambda_3=0.5$	0.5565			0.1201			0.3435		
		REBMIX	w_1	0.7767	-677.644	1386.361	0.8617	-806.0057	1643.084	0.7804	-773.9983	1579.07
			w_2	0.2161			0.1117			0.201		
			$\lambda_1=2.5$	1.1859			1.6244			1.7561		
			$\lambda_2=1.5$	0.2207			0.019			0.2151		
			$\lambda_3=0.5$	4.8793			3.761			3.9039		
	4			$w_1=0.1, w_2=0.3, w_3=0.5, w_4=0.1$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$		
		EM	w_1	0.253	-857.0844	1757.671	0.3209	-913.6899	1870.882	0.3007	-907.4484	1858.399
			w_2	0.1999			0.2335			0.2724		
			w_3	0.3308			0.2321			0.2287		
			$\lambda_1=3.5$	3.0881			3.2067			3.2534		
			$\lambda_2=2.5$	2.1111			2.6606			2.3651		
			$\lambda_3=1$	0.8532			0.5573			0.60466		
			$\lambda_4=0.5$	0.8525			0.5363			0.5604		
		REBMIX	w_1	0.7289	-860.8231	1765.148	0.7117	-922.2272	1887.957	0.7412	-914.366	1872.234
			w_2	0.005			0.0042			0.0043		
			w_3	0.2384			0.135			0.1164		
			$\lambda_1=3.5$	1.9483			1.63			1.6152		
			$\lambda_2=2.5$	8.4491			9.31			9.3022		
			$\lambda_3=1$	0.2928			0.0097			0.0162		
			$\lambda_4=0.5$	4.9933			4.6406			4.5852		

جدول (5): نتائج المحاكاه لتقدير خليط من توزيع بواسون بحجم عينة (1500) مفردة

BIC	LL	التقديرات (اوزان متقاربة)	BIC	LL	التقديرات (اوزان متساوية)	BIC	LL	التقديرات (اوزان متفاوته)	المعلمات الفعلية	خوارزمية التقدير	k	الخليط
$w_1=0.4, w_2=0.6$			$w_1=w_2=0.5$			$w_1=0.2, w_2=0.8$					2	Poisson
4281.937	-2129.999	0.4324	4533.406	-2255.733	0.498	3652.495	-1815.277	0.2169	w_1	EM		
		1.8812			1.966			1.8405	$\lambda_1=2$			
		0.472			0.5091			0.488	$\lambda_2=0.5$			
4288.965	-2133.513	0.6344	4537.679	-2257.87	0.5762	3663.114	-1820.587	0.4512	w_1	REBMIX		
		1.5729			1.8859			1.4056	$\lambda_1=2$			
		0.2282			0.3493			0.2681	$\lambda_2=0.5$			
$w_1=0.25, w_2=0.35, w_3=0.4$			$w_1=w_2=w_3=0.333$			$w_1=0.1, w_2=0.3, w_3=0.6$					3	
4678.44	-2320.937	0.2755	4873.589	-2418.511	0.3306	4088.89	-2026.162	0.3412	w_1	EM		
		0.2428			0.2887			0.2041	w_2			
		2.0719			2.1911			1.8497	$\lambda_1=2.5$			
		1.9797			1.8738			0.5426	$\lambda_2=1.5$			
		0.5614			0.5378			0.5185	$\lambda_3=0.5$			
4680.582	-2322.008	0.5926	4874.871	-2419.152	0.66169	4101.315	-2032.374	0.5913	w_1	REBMIX		
		0.4065			0.3305			0.4052	w_2			
		1.9267			1.9508			1.4741	$\lambda_1=2.5$			
		0.4256			0.4328			0.2155	$\lambda_2=1.5$			
		7.7499			4.7262			5.2376	$\lambda_3=0.5$			
$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.1, w_2=0.3, w_3=0.5, w_4=0.1$					4	
4693.069	-2320.938	0.2290	4888.176	-2418.492	0.2568	4103.516	-2026.162	0.147	w_1	EM		
		0.2863			0.3017			0.195	w_2			
		0.2610			0.2543			0.3398	w_3			
		2.0424			2.1162			1.8586	$\lambda_1=3.5$			
		2.0262			2.1128			1.8399	$\lambda_2=2.5$			
		0.5226			0.8651			0.5287	$\lambda_3=1$			
		0.5205			0.3694			0.5205	$\lambda_4=0.5$			
4695.235	-2322.021	0.5871	4889.906	-2419.357	0.6719	4120.058	-2034.433	0.5962	w_1	REBMIX		
		0.4065			0.318			0.393	w_2			
		0.0009			0.006			0.0038	w_3			
		1.9341			1.9523			1.4752	$\lambda_1=3.5$			
		0.4255			0.3911			0.181	$\lambda_2=2.5$			
		7.7499			4.6854			5.1468	$\lambda_3=1$			
		1.1367			1.4248			1	$\lambda_4=0.5$			

جدول (6): نتائج المحاكاه لتقدير خليط من توزيع بواسون بحجم عينة (2500) مفردة

BIC	LL	التقديرات (اوزان متقاربة)	BIC	LL	التقديرات (اوزان متساوية)	BIC	LL	التقديرات (اوزان متفاوته)	المعلمات الفعلية	خوارزمية التقدير	k	الخليط
$w_1=0.4, w_2=0.6$			$w_1=w_2=0.5$			$w_1=0.2, w_2=0.8$					2	Poisson
7273.2	-3624.864	0.4878	7635.992	-3806.26	0.5525	6207.634	-3092.081	0.3687	w_1	EM		
		1.8563			1.9285			1.525	$\lambda_1=2$			
		0.4494			0.4722			0.4014	$\lambda_2=0.5$			
7275.45	-3625.989	0.4783	7644.389	-3810.459	0.6285	6326.255	-3151.391	0.9996	w_1	REBMIX		
		1.9145			1.8562			0.8129	$\lambda_1=2$			
		0.4216			0.2965			8	$\lambda_2=0.5$			
$w_1=0.25, w_2=0.35, w_3=0.4$			$w_1=w_2=w_3=0.333$			$w_1=0.1, w_2=0.3, w_3=0.6$					3	
7914.094	-3937.487	0.2952	8234.098	-4097.489	0.3576	6918.121	-3439.5	0.217	w_1	EM		
		0.2471			0.2939			0.2362	w_2			
		2.0851			2.1245			1.7376	$\lambda_1=2.5$			
		2.073			2.0627			1.6194	$\lambda_2=1.5$			
		0.5484			0.4545			0.4807	$\lambda_3=0.5$			
7921.307	-3941.094	0.4271	8242.326	-4101.603	0.5034	6921.692	-3441.286	0.3302	w_1	REBMIX		
		0.5727			0.4966			0.6696	w_2			
		2.3525			2.3619			1.9375	$\lambda_1=2.5$			
		0.651			0.6742			0.5696	$\lambda_2=1.5$			
		9			9			8	$\lambda_3=0.5$			
$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.1, w_2=0.3, w_3=0.5, w_4=0.1$					4	
7929.744	-3937.488	0.2464	8249.749	-4097.49	0.3055	6933.771	-3439.501	0.1746	w_1	EM		
		0.2954			0.345			0.2728	w_2			
		0.2426			0.1545			0.2827	w_3			
		2.0808			2.098			1.6915	$\lambda_1=3.5$			
		2.0802			2.098			1.6798	$\lambda_2=2.5$			
		0.5585			0.4959			0.524	$\lambda_3=1$			
		0.5387			0.426			0.4465	$\lambda_4=0.5$			
7937.122	-3941.177	0.4011	8258.184	-4101.708	0.4864	6937.707	-3441.469	0.323	w_1	REBMIX		
		0.5946			0.4966			0.6745	w_2			
		0.0001			7.5e-05			0.0002	w_3			
		2.367			2.3494			1.9671	$\lambda_1=3.5$			
		0.6993			0.6742			0.5732	$\lambda_2=2.5$			
		9			9			8	$\lambda_3=1$			
		3			2.7201			2.4	$\lambda_4=0.5$			

10.النتائج:

توصلت الورقة البحثية من خلال المحاكاة الى النتائج الآتية:

1- ان المعلمات المستخرجة باستخدام خوارزمية REBMIX تعطى نتائج أفضل من حيث جوده الملاءمة ، ويتضح ذلك من خلال الزيادة الملحوظة فى قيمة كل من لوغاريتم الامكان LL ومعيار المعلومات البيزى BIC.

2- بزيادة أحجام العينات تزداد معها قيم كل من لوغاريتم الامكان LL ومعيار المعلومات البيزى BIC وكذلك يظهر التفاوت الشديد بين خوارزمية REBMIX وخوارزمية EM حيث نلاحظ زيادة كبيرة فى قيمتى معيارى المقارنة.

3- أيضا ، بزيادة عدد المكونات تزداد قيم كل من لوغاريتم الامكان LL ومعيار BIC.

4- تزداد دقة التقدير كلما كانت أوزان الخليط متساوية أو متقاربة والعكس صحيح ، فكلما تفاوتت أوزان الخليط كلما انخفضت معها دقة التقدير يتضح ذلك من خلال كل من قيم LL و BIC.

5- من خلال التنفيذ باستخدام حزم برنامج R i386 4.0.3 يلاحظ أن التقدير باستخدام خوارزمية REBMIX يعطى نتائج اسرع من استخدام خوارزمية EM.

6- عند استخدام اوزان متقاربة أو متساوية فإن التقديرات تكون أقرب الى المعلمة الحقيقية من استخدام اوزان متفاوته وذلك لكل من خوارزميتى EM و REBMIX.

المراجع:

……. (2021, january 4). *Bayesian information criterion*. Retrieved from https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/WIKIPEDI/W120607B.pdf.

Chris , F., & Raftery, A. (2017, September 6). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association, 5*(1).

Claeskens, G., & Hjort, N. (2008). *Model Selection and Model Averaging.* New York, USA: Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B, 39*(1).

Filho, I. (2008). *Mixture Models for the Analysis of Gene Expression: Integration of Multiple Experiments and Cluster Validation.* Berlin, Germany: Department of Mathematics and Computer Science, Free University of Berlin, Germany.

Fraley, C., & Raftery, A. (2002, jun). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association, 97*(458).

Frempong, N., & et al. (2017, September 6). Fitting finite mixtures of generalized linear regressions on motor insurance claims. *International Journal of Statistics and Actuarial Science*(5).

Ghojogh, B., & et al. (2019, 12 14). Fitting A Mixture Distribution to Data: Tutorial. Waterloo, canada.

Grün, B., & Leisch, F. (2007). *Finite mixtures of generalized linear regression models.* Munich, Germany: Department of Statistics University of Munich.

Klugman, S., & et al. (2012). *Loss Models From Data to Decisions Fourth Edition.* New Jersey,USA: John Wiley & Sons.

Nagode, M. (2013). *rebmix: An R Package for Continuous and Discrete Finite Mixture Models.* http://www2.uaem.mx/r-mirror/web/packages/rebmix/vignettes/rebmix.pdf.

Nagode, M., & Fajdiga, M. (2011). The REBMIX Algorithm and the Univariate Finite Mixture Estimation. *Communications in Statistics - Theory and Methods, 40*(11).

Nagode, M., & Fajdiga, M. (2011). The REBMIX Algorithm and the Univariate Finite Mixture Estimation. *Communications in Statistics - Theory and Methods, 40*(11).

Ng, S., & et al. (2019). *Mixture Modelling for Medical and Health Sciences.* FL,USA: Taylor & Francis Group.

Pani´c, B., & et al. (2020, March). Improved Initialization of the EM Algorithm for Mixture Model Parameter Estimation. *Mathematics, 8*(4).

Panić, B., & at el. (2020). Gaussian Mixture Model Based Classification Revisited: Application to the Bearing Fault Classification. *Journal of Mechanical Engineering, 66*(4).

Peter , A., & Jones, J. (2003, May). Finite Mixture Distributions, Sequential Likelihood and The EM Algorithm. *Econometrica, 71*(3).

Yen, P., Ismail, M., & Hamzah, F. (2014). finite mixture model a maximum likelihood estimation approach on time series data. *Statistics and Operational Research International Conference*, *1613.*