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

محاكاة توزيعات وايبل المختلطة المحدودة لنمذجة حجم المطالبة التأمينية

Simulation of finite mixture Weibull distributions for modeling insurance claim

الباحث/ احمد رمضان محمد المتولى
المدرس المساعد بقسم الاحصاء والرياضيات والتأمين
بكلية 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:

تعتبر التوزيعات المستمرة من التوزيعات الهامة المستخدمة فى التقدير والتنبؤ بحجم الخسائر المالية المتوقعة. من هنا ، تسعى هذه الورقة البحثية الى عرض طريقتى تقدير لمجموعة من التوزيعات المختلطة المحدودة والمتطابقة والتى يمكن استخدامها فى نمذجة حجم المطالبة التأمينية المتوقعة ، باستخدام أسلوب المحاكاه وخوارزميتن هما: خوارزمية تعظيم التوقع Expectation Maximization (EM) Algorithm وخوارزمية تقدير بايز المحسن التقريبى The Rough-Enhanced-Bayes Mixture (REBMIX) algorithm. وباستخدام توزيع مستمر هو توزيع وايبل بمعلمتين 2-Parameter Weibull ، اعتمادا على أوزان خليط مختلفة وعدد مكونات مختلفة وأحجام عينات مختلفة.

الكلمات المفتاحية : توزيع وايبل ، التوزيعات المختلطة المحدودة ، خوارزمية تعظيم التوقع ، خوارزمية تقدير بايز المحسن التقريبى ، المحاكاة.

Abstract:

Continuous distributions are one of the important distributions used in estimating and predicting the expected financial losses. From here, this paper seeks to present two estimation methods for a set of identical finite mixture distributions that can be used to modeling the claim severity, using the simulation method and two algorithms: Expectation Maximization (EM) Algorithm and The Rough-Enhanced-Bayes Mixture (REBMIX) algorithm. Note that, the distribution used is a 2-Parameter Weibull distribution, depending on different mixture weights, number of different components, and different sample sizes.

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

1. المقدمة Introduction:

توزيعات الخسارة طريقة رياضية لنمذجة المطالبات الفردية. هذه المطالبات عادتا ما تكون غير متجانسة لكن لا يمكن فصل تأثير عدم التجانس منها. من هنا فإن استخدام التوزيعات المختلطة المحدودة توفر أسلوب احصائى يمكن الاعتماد عليه لوصف عدم التجانس غير الملاحظ فى تلك البيانات. فى هذه الورقة البحثية يتم عرض بعض التوزيعات الاحتمالية المرتبطة بحجم الخسارة وكيفية تركيبها أو تهذيبها Fitted باستخدام بيانات محاكاة لحجم المطالبات المشاهدة.

من هنا ، تسعى هذه الورقة الى محاولة البحث عن طريقة جديدة لتقدير معلمات توزيع حجم المطالبات بشكل يراعى عدم التجانس غير الملاحظ فى بيانات المجتمع. ولعل الاجراء الأكثر شيوعا للوصول الى افضل تقدير لمعلمات التوزيع المختلط المحدود هو استخدام خوارزمية EM_algorithm Expectation Maximization وهو اجراء يهدف الى الحصول على اقصى قيمة للوغاريتم الامكان المتوقع.

بالاضافة الى خوارزمية تعظيم التوقع EM توجد خوارزمية أخرى تسمى خوارزمية تقدير بايز المحسن التقريبى The Rough-Enhanced-Bayes Mixture (REBMIX) algorithm والتى ظهرت على يد كل من Nagode and Fajdiga (1998) بهدف تجنب بعض العيوب التى ظهرت مع استخدام طريقة EM (Nagode, 2013). وباختصار ، فإن المشكلة تتمثل فى تصميم محاكاة لمحاولة الوصول الى أفضل طريقة تقدير من بين الطريقتين المعروضتين فى الورقة للتعبير عن حجم المطالبات التأمينية لشركات التأمين على الممتلكات من خلال مقارنة جودة التقديرات.

2. توزيع وايبل Weibull Distribution

ويعتبر توزيع وايبل احد اشهر توزيعات مدى الحياة والذى اقترح من قبل العالم الفيزيائى السويدى Waloddi Weibull عام 1939م (Lai & Xie, 2006). وللتوضيح ، وتكون دالة كثافة الاحتمال للمتغير العشوائى X والذى يتبع توزيع وايبل ، على الصورة:

$$f(x)=\begin{cases}\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-\left(x/\beta\right)^{\alpha}} & , \quad x\ge 0\\ 0 \quad , \quad x<0\end{cases}$$

حيث α تسمى معلمة الشكل ، و β معلمة القياس حيث $\alpha,\beta>0$.

3. التوزيعات المختلطة Mixture Distributions

بفرض أن $X = (X_1, X_2, \dots, X_i, \dots, X_L)$ ، تمثل متغير عشوائى مستمر من البعد L ، كما أن $x = (x_1, x_2, \dots, x_i, \dots, x_L)$ تمثل قيم المشاهدات لـ X. وبفرض أن دالة الكثافة الاحتمالية (pdf) للنموذج المختلط يمكن التعبير عنها كتركيبة محدبة من عدد K من دوال الكثافة الاحتمالية الفردية (Zhang & Huang, 2015) (Volodymyr & Ranjan, 2010) ، على الصورة:

$$f(x;\theta) = \sum_{k=1}^{K} w_k g_k(x;\theta_k) \quad , \quad w \geq 0 \ , \ \ i \in \{1,2,...,k\} \ , \ \ \sum_{k=1}^{K} w_i = 1$$

حيث تمثل $g_k(x;\theta_k)$ دالة الكثافة الاحتمالية للمكون رقم k ، و w_k نسبة المكون ، كما ان $\theta = (w_1, w_2, ..., w_k, \theta_1, \theta_2, ..., \theta_k)$ تعبر عن مجموعة المعلمات ، و $f(x;\theta)$ تمثل دالة الكثافة الاحتمالية للخليط الناتج من مجموعة من الكثافات الاحتمالية الفردية $g_k(x;\theta_k)$.

4. تقدير نماذج الخليط المحدود Estimation in Finite Mixture Models

فى هذه الدراسة سوف نتناول خوارزميتى تقدير هما: خوارزمية تعظيم التوقع EM Algorithm وخوارزمية تقدير باير المحسن التقريبى هى REBMIX Algorithm.

4.1 خوارزمية تعظيم التوقع EM Algorithm

تمثل خوارزمية تعظيم التوقع (EM) طريقة عامة لإيجاد مقدرات الإمكان الأعظم فى حال وجود قيم مفقودة Missing Values أو متغيرات كامنة Latent Variable (Filho, 2008). حيث فى كثير من الأحيان نجد أن البيانات لا تعطينا صورة كاملة حول الظاهرة وبالتالى تقوم EM بتعظيم قيمة الاحتمال بشكل تكرارى عوضاً عن استخدام القيم المشاهدة للوغاريتم دالة الإمكان. من خلال خطوتين **الأولى E-Step**: وفيها يتم الحصول على القيمة المتوقعة لمعادلة الامكان حيث يتم اخذ توقع لوغاريتم الامكان بهدف الحصول على تقديرات متوسطة للقيم المفقودة $D^{(Miss)}$. وبافتراض أن $Q(\theta)$ ترمز الى توقع دالة الامكان الخاص بقيم $D^{(Miss)}$ حيث:

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

مما سبق يتبين أن هذا التوقع مشروط بكل من $D^{(obs)},\theta$ حيث تعبر $D^{(obs)}$ عن القيم المشاهدة لذا يتم التعامل معها هنا كثوابت وليست متغيرات.

اما المرحلة أو الخطوة الثانية M_Step: فتعتمد على استخدام طريقة MLE ، حيث يتم التعويض عن لوغاريتم الامكان بالتوقع ، على الصورة أى:

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

بفرض وجود عدد k من مكونات الخليط من التوزيعات ذات الكثافات الاحتمالية $g_1(x;\theta_1),...,g_k(x;\theta_k)$.

وبافتراض أن التوزيع احادى المتغير Univariate distribution وعليه فإن دالة كثافة الاحتمال للخليط هى (Hesse, Holtackers, & Heskes, 2006):

$$f(x;\theta)=\sum_{k=1}^{K} w_k g_k(x;\theta_k)$$

وباستخدام خوارزمية EM لتقدير المعلمات وفقاً لأكبر توقع محلى متوقع على مرحلتين:

بالنسبة للخطوة الأولى ، يتم حساب التوقع وفقا للمعادلة:

$$Q(\theta_1,...,\theta_K)=\sum_{i=1}^{n}\left[\sum_{k=1}^{K}E\left[\Delta_{i,k}\middle|X,\theta_1,...,\theta_K\right]\times\log\left(w_k g_k(x_i;\theta_k)\right)\right]$$

وحيث أن $\Delta_{i,k}$ تساوى "1" فى حال كانت المشاهدة رقم i تنتمى للمكون k و "0" اذا كانت لا تنتمى ، فإن:

$$E\left[\Delta_{i,k}\middle|X,\theta_1,...,\theta_K\right]=0\times P\left(\Delta_{i,k}=0\middle|X,\theta_1,...,\theta_K\right)+1\times P\left(\Delta_{i,k}=1\middle|X,\theta_1,...,\theta_K\right)=P\left(\Delta_{i,k}=1\middle|X,\theta_1,...,\theta_K\right)$$

وباستخدام قاعدة بايز للحصول على الاحتمال البعدى ، نجد أن:

$$P\left(\Delta_{i,k}=1\middle|X,\theta_1,...,\theta_K\right)=\frac{P\left(X,\theta_1,...,\theta_K,\Delta_{i,k}=1\right)}{P\left(X;\theta_1,...,\theta_K\right)}=\frac{P\left(X,\theta_1,...,\theta_K|\Delta_{i,k}=1\right)P\left(\Delta_{i,k}=1\right)}{\sum_{k'=1}^{K}P\left(X,\theta_1,...,\theta_K|\Delta_{i,k'}=1\right)P\left(\Delta_{i,k'}=1\right)}$$

حيث قيمة الاحتمال الهامشى The marginal probability فى المقام على الصورة:

$$P(X;\theta_1,...,\theta_K)=\sum_{k'=1}^{K} w_{k'} g_{k'}(x_i,\theta_{k'})$$

وبافتراض أن المقدار $E\left[\Delta_{i,k}\middle|X,\theta_1,...,\theta_K\right]$ يعبر عن Responsibility للقيم x_i ويرمز له بالرمز $\hat{\gamma}_{i,k}$ (Ng & et al, 2019) ، حيث:

$$\hat{\gamma}_{i,k}=\frac{\hat{w}_k g_k\left(x_i,\hat{\theta}_k\right)}{g\left(x_i,\hat{\theta}_k\right)}=\frac{\hat{w}_k g_k\left(x_i,\hat{\theta}_k\right)}{\sum_{k'=1}^{K}\hat{w}_{k'} g_{k'}\left(x_i,\hat{\theta}_{k'}\right)}$$

من هنا يمكن إعادة حساب التوقع فى المعادلة على الصورة:

$$Q(\theta_1,...,\theta_K)=\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\hat{\gamma}_{i,k}\log\left(w_k g_k(x_i;\theta_k)\right)\right]$$

وبتوزيع اللوغاريتم على المقدار $w_k g_k(x_i;\theta_k)$ يمكن اعادة تبسيط المعادلة السابقة على الصورة:

$$Q(\theta_1,...,\theta_K)=\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\hat{\gamma}_{i,k}\log w_k+\hat{\gamma}_{i,k}\log\left(g_k(x_i;\theta_k)\right)\right]$$

اما الخطوة الثانية والتى يتم فيها محاولة تعظيم خوارزمية التوقع للحصول على أقصى قيمة محلية ممكنه (Ghojogh & et al., 2019) ، من خلال العلاقة:

$$\hat{\theta}_k,\hat{w}_k=\arg\max_{\theta_k,w_k}\ Q\left(\theta_1,...,\theta_K,w_1,...,w_k\right)\quad,\quad\sum_{k=1}^{K}w_k=1$$

وباضافة معامل لاجرانج ، يمكن اعادة كتابة لوغاريتم دالة الامكان ليصبح على النحو التالى:

$$LL\left(\theta_1,...,\theta_K,w_1,...,w_k\right)=Q\left(\theta_1,...,\theta_K,w_1,...,w_k\right)-\alpha\left(\sum_{k=1}^{K}w_k-1\right)=\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\hat{\gamma}_{i,k}\log w_k+\hat{\gamma}_{i,k}\log g_k\left(x_i;\theta_k\right)\right]-\alpha\left(\sum_{k=1}^{K}w_k-1\right)$$

وباستخدام المشتقة الأولى ومساواتها بالصفر يمكن ايجاد معلمة التوزيع للمكون k:

$$\frac{\partial LL}{\partial\theta_k}=\sum_{i=1}^{n}\frac{\hat{\gamma}_{i,k}}{g_k\left(x_i;\theta_k\right)}.\frac{\partial g_k\left(x_i;\theta_k\right)}{\partial\theta_k}\overset{set}{=}0$$

بالمثل يمكن ايجاد معلمة الخلط للمكون رقم k على النحو التالى:

$$\frac{\partial LL}{\partial w_k}=\sum_{i=1}^{n}\frac{\hat{\gamma}_{i,k}}{w_k}-\alpha\overset{set}{=}0\Rightarrow w_k=\frac{1}{\alpha}\sum_{i=1}^{n}\hat{\gamma}_{i,k}$$

كذلك فإن:

$$\frac{\partial LL}{\partial\alpha}=\sum_{k=1}^{K}w_k-1\overset{set}{=}0\Rightarrow\sum_{i=1}^{n}w_k=1$$

$$\sum_{k=1}^{K}\frac{1}{\alpha}\sum_{i=1}^{n}\hat{\gamma}_{i,k}=1\Rightarrow\alpha=\sum_{i=1}^{n}\sum_{k=1}^{K}\hat{\gamma}_{i,k}$$

وبالتعويض فى المعادلة نجد أن:

$$\hat{w}_k=\sum_{i=1}^{n}\hat{\gamma}_{i,k}\Big/\sum_{i=1}^{n}\sum_{k=1}^{K}\hat{\gamma}_{i,k}$$

وبحل المعادلات السابقة بصورة تكرارية يمكن الحصول على تقديرات للمعلمات $\hat{\theta}_k,\hat{w}_k$.

4.2 خوارزمية REBMIX:

ان المنهج الكلاسيكى فى بناء الخليط المحدود يعتمد على استخدام خوارزمية EM فى حل معادلة الامكان ، وبالتالى يمكن تقدير جميع المعلمات غير المعروفة فى نموذج الخليط بصورة تكرارية وبالتالى تضمن خوارزمية EM ان قيمة الامكان تزداد لكن هذه الزيادة تصل الى ان تكون على وتيرة واحدة أى لا يتغير تقريبا (Nagode & Fajdiga, 2011). ان خوارزمية REBMIX تحاول معالجة مشكلة الرتابة التى تظهر مع استخدام خوارزمية EM فى التقدير وبالتالى فهى تمثل اجراء عددى تكرارى يحتوى على سبع خطوات يجب اتباعها عند بناء النموذج هذه الخطوات تتمثل فى:

1) المعالجة المسبقة للمشاهدات Preprocessing of observations

الخطوة الأولى من خطوات بناء الخوارزمية وذلك عن طريق اجراء تقدير للكثافة التجريبية لمجموعة البيانات (Panić & at el., 2020). وباستخدام المدرج التكرارى يتم تجزئة عدد المشاهدات الإجمالى n والذى يمثل اجمالى المشاهدات العددية المستقلة الى عدد محدود من الأجزاء bins تمثل كل منها عمود من أعمدة المدرج التكرارى تتميز بأنها غير متداخلة ومتساوية الحجم وموزعة بانتظام (Nagode, 2013). وبافتراض ان متوسطات هذه المجموعات bins y_j يمكن حسابها عن طريق المعادلة (Nagode & Fajdiga, 2011):

$$y_j = y_0 + \text{'An arbitrary integer'} \times \text{h}$$

ان الكسر Fraction فى المشاهدات C_i والذى يقع فى الحجم R_i يخرج من العد حيث $i = 1, 2, ..., c$ ، و y_0 يرمز الى الأصل الحكمى Arbitrary origin ويبين c العدد الاجمالى للأعمدة داخل المدرج التكرارى.

وتتحدد طول فئة المدرج التكرارى بالمعادلة (Nagode , 2015):

$$h = {(y_{\max} - y_{\min})}\big/{c}$$

ويتم ايجاد الأصل y_0 من خلال المعادلة:

$$y_0 = y_{\min} + {h}\big/{2}$$

2) اكتشاف النمط العام Global Mode Detection

يعتمد اكتشاف النمط العام أساساً لتقدير معلمة المكون التقريبية (Panić & at el., 2020).
ويتطابق النمط العام لـ y_m ، نجد أن الكثافة f_{kj} التجريبية تأخذ اقصى قيمة على الصورة:

$$m = \arg\max\ f_{kj} \to (y_m, f_{km})$$

وبالوصل الى الحد الأقصى المحلى والذى يحدد النمط العام ، عندها يتم وضع المشاهدات فى مجموعات داخل المدرج التكرارى (Nagode, 2013) ، وبالتالى نجد أن:

$$f_{kj} = \frac{C_{kj}}{n_k} \frac{1}{h_j} \qquad j = 1, 2, ..., c$$

حيث يتم تهيئة التكرارات C_{kj} بصورة مبدأية على C_j ويكون عدد المشاهدات فى الفئة k (Nagode , 2015)، على النحو التالى:

$$n_k = \sum_{j=1}^{c} C_{kj}$$

3) تجميع المشاهدات Clustering of Observations

يعتبر تجميع المشاهدات بمثابة اجراء تكرارى لتحديد المشاهدات التى تنتمى الى مكون معين وليكن k. وحيث انه سوف يتم التركيز هنا حول النماذج وحيدة المتغيرات فإن الخوارزمية REBMIX تفترض فى البداية ان الخليط يتكون من مكون واحد اى أن $k = 1$ وبالتالى فإن $n_1 = n$ ، $c_{kj} = c_j$ حيث $j = 1, 2, ..., c$ والتكرارات المتبقية $r_j = 0$ لجميع قيم j. ويمكن حساب الانحرافات بين c_{kj} وتكرارات المكون التنبؤية باستخدام منهج المدرج التكرارى على الصورة:

$$e_{kj} = c_{kj} - n_k f\left(y_j \middle| \theta_k\right) h_j$$

وتحسب الانحرافات النسبية للمشاهدات على الصورة:

$$\varepsilon_{kj} = {e_{kj}}/{c_{kj}}$$

ويمكن فصل كل من مجموع الانحرافات الموجبة e_{kp} ومجموع الانحرافات السالبة e_{kn} كما يلى:

$$e_{kp} = \sum_{j=1, e_{kj}>0}^{c} e_{kj} , e_{kn} = \sum_{j=1, e_{kj}<0}^{c} e_{kj}$$

وبنفس الطريقة ، يمكن حساب الانحرافات النسبية الموجبه على النحو التالى:

$$D_k = e_{kp}/n_k$$

حيث D_k تنحصر نتائجها بين الصفر والواحد الصحيح.

4) تقدير معلمة المكون التقريبية Rough Component Parameter Estimation

فى هذه الخطوة يتم استخراج العناصر المحيطة للنمط العام أولاً ، والتى من المفترض أن يكون هناك مكون واحد على الأقل فى المنطقة المجاورة له (Nagode & Fajdiga, 2006). وبالتالى فإنه يمكن ضمان الاستخلاص الأمثل للمشاهدات التى تنتمى الى المكون رقم k من خلال القيود التى تمنع المكون من التدفق بعيدا عن النمط العام ، والتى من المفترض أن يكون هناك مكون واحد على الأقل فى المنطقة المجاورة له. ويضمن القيد الأول مساواة كثافات الاحتمال عند y_m ، حيث:

$$f_{km} = f\left(y = y_m | \theta_k\right) = f_{k\max}$$

اما القيد الثانى فيجعل النمط العام لكثافة المكون يتطابق مع y_m من خلال ايجاد المشتقة الأولى لـ $f\left(y = y_m | \theta_k\right)$ ومساواتها بالصفر على النحو التالى:

$$\frac{\partial f\left(y = y_m | \theta_k\right)}{\partial y} = 0$$

ويمكن ايجاد المعلمات لكل من التوزيعات: الطبيعى والطبيعى اللوغاريتمى ووايبل احادية المتغيرات (Nagode & Fajdiga, 2006)، على النحو التالى.

$$\mu_k = y_m \quad , \quad \sigma_k = 1/ f_{km}\sqrt{2\pi}$$

بالمثل ، نجد أن معلمات المكون الطبيعى اللوغاريتمى تكون على الصورة:

$$\mu_k = \ln(y_m) + \sigma_k^2 \quad , \quad f(\sigma_k) = f_{km} y_m \sigma_k \sqrt{2\pi} - e^{-\frac{\sigma_k^2}{2}} = 0$$

واخيراً ، فمعلمات مكون وايبل تكون على الصورة:

$$f(\alpha_k) = f_{km} y_m - (\alpha_k - 1) e^{-\frac{\alpha_k - 1}{\alpha_k}} = 0 \quad , \quad \beta_k = y_m \left(\alpha_k - 1/\alpha_k\right)^{-\frac{1}{\alpha_k}} + \sigma_k^2 \quad , \quad \alpha_k > 1$$

5) تقدير معلمة المكون المحسنة Enhanced Component Parameter Estimation

للحصول على معلمات المكون المحسنة ، يتم استخدام طريقة الامكان الأعظم. وبتطبيق مدخل المدرج التكرارى ، يمكن ايجاد معلمات المكون المحسنة والذى يتبع التوزيع الطبيعى على الصورة:

$$\mu_k = \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} y_i \qquad , \qquad \sigma_k^2 = \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} y_j^2 - \mu_k^2$$

بالمثل ، يمكن تحسين معلمة المكون الذى يتبع التوزيع الطبيعى اللوغاريتمى على الصورة:

$$\mu_k = \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} \ln(y_m) \qquad , \qquad \sigma_k^2 = \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} \ln\left(y_j^2\right)^2 - \mu_k^2$$

وأخيرا ، يمكن تحسين معلمة المكون الذى يتبع توزيع وايبل على الصورة:

$$f(\alpha_k) = \frac{1}{\alpha_k} + \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} \ln(y_j) - \frac{\sum_{j=1}^{c} C_{kj} y_j^{\alpha_k} \ln(y_j)}{\sum_{j=1}^{c} C_{kj} y_j^{\alpha_k}} = 0$$

$$\beta_k^{\alpha_k} = \frac{1}{n_k}\sum_{j=1}^{c} C_{kj} y_j^{\alpha_k}$$

6) حساب العزم الأول والثانى First and second Moment Calculation

باستخدام العزم الأول والثانى يمكن ايجاد متوسط وتباين المكون الذى يتبع التوزيع الطبيعى (Nagode, 2013) ، على النحو التالى:

$$m_k = \mu_k \qquad , \qquad Var_k = \sigma_k^2 + \mu_k^2$$

بالمثل ، يمكن ايجاد متوسط وتباين المكون الذى يتبع التوزيع الطبيعى اللوغاريتمى على الصورة:

$$m_k = e^{\mu_k + \frac{\sigma_k^2}{2}} \qquad , \qquad \mathrm{Var}_k = e^{2\mu_k + 2\sigma_k^2}$$

7) تصنيف بايز للمشاهدات المتبقية Bayes Classification of the Remaining Observations

بزيادة عدد المكونات ، ينخفض عدد المشاهدات المتبقية. وعندما يصل الوزن النسبى للمكون الى اقل وزن نسبى (Nagode & Fajdiga, 2011) ، أى:

$$w_k \le w_{\min}$$

يتم افتراض أن المشاهدات المتبقية تنتمى الى الفئة الحالية ولا يتم تكوين فئات جديدة.

ايضا فهناك لاقة ارتباط بين اقل وزن نسبى و $D_{\min}$. فكلما زاد مؤشر المكون رقم K فإن كل من عدد المكونات ومجموع الانحرافات النسبية الموجبة يزداد ايضاً.

$$D = \sum_{k=1}^{K} w_k D_k \geq K D_{\min}$$

واذا وصل الوزن النسبى للمكون الى اجمالى الانحرافات النسبية بمقدار $2kD_{\min}$ تتوقف الخوارزمية REBMIX.

$$w_{\min} = 2kD_{\min}$$

ويتم تصنيف المشاهدات المتبقية اعتمادا على قاعدة بايز ، كما يلى:

$$w_k = w_k + \frac{C_{kj}}{n} \quad , \quad m_1 = m_1 + \frac{C_{kj}\left(y_j - m_k\right)}{nw_k}$$

$$Var_k = Var_k + \frac{C_{kj}\left(y_j^2 - Var_k\right)}{nw_k}$$

حيث يتم اضافة C_{kj} الى الفئة k والوزن النسبى للمكون ، ويتم اعادة حساب المتوسط والتباين للمكون. بمجرد معالجة جميع المجوعات والتى يرمز لها بالرمز C او جميع المشاهدات n ، يتم الحصول على معلمات الخليط عن طريق اعادة حساب العزم الأول والثانى لمكونات الخليط.

5. المحاكاة:

فى هذه الدراسة تم استخدام كل من لوغاريتم الامكان 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) كلما كانت قيمته اعلى دل ذلك على جودة التقديرات، كذلك سوف يتم استخدام خوارزميتى EM و REBMIX ومن ثم مقارنة نتائج التقدير وباستخدام حزم برنامج R باصدار R i386 4.0.3.

6.1 تصميم المحاكاه:

تمت عملية المحاكاة باستخدام بيانات افتراضية من أحجام مختلفة من العينات وهى 500 و 1500 و 2500 مفردة بنسب تمثيل مختلفة. ويتم تكوين توزيعات مختلطة محدودة متماثلة Identical Finite Mixtue Distributions من عدد من المكونات k مفترضة للدراسة ، حيث k=2,3,4. ايضا يمكن عرض توزيعات الخليط التى سيتم بنائها ومحاكاتها وفقا للجدول:

جدول (1): توزيعات الخليط المولدة لعملية المحاكاه بالنسبة لتوزيع وايبل

عدد المكونات	المعلمات
2	$w, \alpha_{w1}, \beta_{w1}, \alpha_{w2}, \beta_{w2}$
3	$w_1, w_2, \alpha_{w1}, \beta_{w1}, \alpha_{w2}, \beta_{w2}, \alpha_{w3}, \beta_{w3}$
4	$w_1, w_2, w_3, \alpha_{w1}, \beta_{w1}, \alpha_{w2}, \beta_{w2}, \alpha_{w3}, \beta_{w3}, \alpha_{w4}, \beta_{w4}$

أيضا تتم عملية المحاكاه مع افتراض أن اوزان للخليط على ثلاث حالات وفقا للجدول التالى:

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 فقد تم افتراض قيماً تقريبية للمعلمات الحقيقية من خلال النظر الى طبيعة بيانات المطالبة التأمينية للبيانات الفعلية المتاحة ، وفقا للجدول التالى:

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

توزيعات الخليط	معلمات المكونات
Weibull(2)	$\alpha_{w1} = 0.72738, \beta_{w1} = 409.1349, \alpha_{w2} = 0.60222, \beta_{w2} = 1335.6705$
Weibull (3)	$\alpha_{w1} = 0.5426926, \beta_{w1} = 86.28743, \alpha_{w2} = 0.7855334, \beta_{w2} = 609.64774, \alpha_{w3} = 0.5316595, \beta_{w3} = 1001.696$
Weibull(4)	$\alpha_{w1} = 0.511479, \beta_{w1} = 156.2808, \alpha_{w2} = 0.77344, \beta_{w2} = 559.484, \alpha_{w3} = 0.52639, \beta_{w3} = 656.8403, \alpha_{w4} = 0.62661, \beta_{w4} = 1683.936$

6.2 نتائج المحاكاه لتوزيعات وايبل الممثلة لحجم المطالبة:

تتمثل توزيعات حجم المطالبة فى التوزيعات شديدة الالتواء ناحية اليمين والتى من اشهرها توزيع وايبل. وقد تم اثناء المحاكاه فرض قيد على عدد التكرارات بحيث 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 اعتمادا على مدخل المدرج التكرارى ، ويمكن عرض النتائج وفقا للقيم المرصودة بالجداول التالية:

جدول (4): نتائج المحاكاه لتقدير خليط من توزيع وايبل 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
weibull	2			$w_1 = 0.8, w_2 = 0.2$			$w_1 = w_2 = 0.5$			$w_1 = 0.6, w_2 = 0.4$		
		EM	w_1	**0.704987**	**−3690.982**	**7413.038**	**0.21391**	**−3818.373**	**7667.818**	**0.20051**	−3816.375	7663.823
			$\alpha_{w1} = 0.72738$	**0.75983**			**0.78609**			**1.28552**		
			$\beta_{w1} = 409.1349$	**291.0358**			**1.17756**			**359.6312**		
			$\alpha_{w2} = 0.60222$	**0.724874**			**345.4097**			**0.56816**		
			$\beta_{w2} = 1335.6705$	**1500.235**			**766.2729**			**745.964**		
		REBMIX	w_1	**0.8746**	**−370.597**	**7552.266**	**0.9112**	**−3939.075**	**7909.222**	**0.910644**	−3936.678	7904.429
			$\alpha_{w1} = 0.72738$	**1.176343**			**0.1088**			**1.0877**		
			$\beta_{w1} = 409.1349$	**427.3416**			**827.508**			**822.762**		
			$\alpha_{w2} = 0.60222$	**1.528405**			**1.03941**			**1.03638**		
			$\beta_{w2} = 1335.6705$	**4174.087**			**5240.71**			**5214.196**		
	3			$w_1 = 0.6, w_2 = 0.3, w_3 = 0.1$			$w_1 = w_2 = w_3 = 0.333$			$w_1 = 0.4, w_2 = 0.35, w_3 = 0.25$		

BIC	LL	التقديرات (اوزان متقاربة)	BIC	LL	التقديرات (اوزان متساوية)	BIC	LL	التقديرات (اوزان متفاوته)	المعلمات الفعلية	خوارزمية التقدير	K	الخليط
6808.749	−3379.516	0.25983	6987.111	−3468.697	0.30428	6538.795	−3244.539	0.39543	w_1	EM		
		0.36939			0.49753			0.43523	w_2			
		0.48393			0.43373			0.50767	$\alpha_{w1}=0.5426926$			
		15.78208			78.3086			29.4079	$\beta_{w1}=86.28743$			
		0.84992			0.61829			0.81224	$\alpha_{w2}=0.7855334$			
		360.6678			494.1171			512.442	$\beta_{w2}=609.64774$			
		0.54188			0.42451			0.43696	$\alpha_{w3}=0.5316595$			
		868.455			768.722			589.981	$\beta_{w3}=1001.696$			
7921.197	−3935.74	0.53188	8194.755	−4072.519	0.51014	8184.564	−4067.424	0.49703	w_1	REBMIX		
		0.05072			0.05028			0.04115	w_2			
		1.18313			1.14737			1.24447	$\alpha_{w1}=0.5426926$			
		975.591			1138.196			873.430	$\beta_{w1}=86.28743$			
		0.67808			2.51753			2.82604	$\alpha_{w2}=0.7855334$			
		3763.205			12734.66			7418.648	$\beta_{w2}=609.64774$			
		1.70386			1.70386			1.70386	$\alpha_{w2}=0.5316595$			
		670.456			768.7402			655.2182	$\beta_{w2}=1001.696$			
$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.1, w_2=0.5, w_3=0.3, w_4=0.1$						
7403.422	−3667.531	0.2783	7356.897	−3644.268	0.29642	7503.713	−3717.676	0.182522	w_1	EM	4	
		0.30292			0.32052			0.469796	w_2			
		0.27346			0.19091			0.074509	w_3			
		0.555786			0.44685			0.570702	$\alpha_{w1}=0.511479$			
		33.34342			148.5197			32.410498	$\beta_{w1}=156.2808$			
		1.30302			0.642517			1.258008	$\alpha_{w2}=0.77344$			
		362.8535			414.4967			426.83012	$\beta_{w2}=559.484$			
		0.9543			0.563712			5.860645	$\alpha_{w3}=0.52639$			
		1566.8228			737.433			2331.1297	$\beta_{w3}=656.8403$			
		0.53844			0.54498			0.568945	$\alpha_{w4}=0.62661$			
		1701.6586			1498.702			1328.35796	$\beta_{w4}=1683.936$			
8190.342	−4060.99	0.48906	7889.355	−3910.497	0.474068	7805.688	−3868.664	0.941526	w_1	REBMIX		
		0.05028			0.02787			0.0281856	w_2			
		0.02689			0.00373			0.004005	w_3			
		1.23585			1.258504			1.067101	$\alpha_{w1}=0.511479$			
		1008.814			1113.868			746.9577	$\beta_{w1}=156.2808$			
		2.37839			3.579439			10.14989	$\alpha_{w2}=0.77344$			
		7353.172			10009.04			8442.077	$\beta_{w2}=559.484$			
		9.0626			58.08894			25.7953	$\alpha_{w3}=0.52639$			
		33489.67			32997.57			12348.25	$\beta_{w3}=656.8403$			
		1.703863			1.001419			8.61342	$\alpha_{w4}=0.62661$			
		578.814			702.9292			33514.35	$\beta_{w4}=1683.936$			

جدول (5): نتائج المحاكاه لتقدير خليط من توزيع وايبل 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
$w_1=0.6, w_2=0.4$			$w_1=w_2=0.5$			$w_1=0.8, w_2=0.2$					2	Weibull
22791.047	−11377.241	0.55317	23228.223	−11595.829	0.48071	22104.278	−11033.856	0.85671	w_1	EM		
		0.69381			0.66577			0.66061	$\alpha_{w1}=0.72738$			
		374.939			366.225			401.154	$\beta_{w1}=409.1349$			
		0.50494			0.518701			0.48294	$\alpha_{w2}=0.60222$			
		1080.784			1243.275			1506.704	$\beta_{w2}=1335.6705$			
23750.38	−11856.91	0.92599	24212.03	−12087.73	0.90656	23842.34	−11902.89	0.96226	w_1	REBMIX		
		1.118423			1.114399			1.14091	$\alpha_{w1}=0.72738$			
		772.4333			815.7947			733.454	$\beta_{w1}=409.1349$			
		1.036263			1.141252			3.21113	$\alpha_{w2}=0.60222$			
		9066.776			9730.179			10074.45	$\beta_{w2}=1335.6705$			
$w_1=0.4, w_2=0.35, w_3=0.25$			$w_1=w_2=w_3=0.333$			$w_1=0.6, w_2=0.3, w_3=0.1$					3	
20666.7	−10304.097	0.06008	21151.447	−10546.471	0.120356	19000.993	−9471.244	0.64854	w_1	EM		
		0.63967			0.595902			0.328308	w_2			
		0.937253			0.59893			0.46014	$\alpha_{w1}=0.5426926$			
		9.730003			25.0848			5937188	$\beta_{w1}=86.28743$			
		0.571487			0.57837			0.75923	$\alpha_{w2}=0.7855334$			
		369.6408			474.845			696.6726	$\beta_{w2}=609.64774$			
		0.348071			0.35988			0.596385	$\alpha_{w3}=0.5316595$			
		350.67748			494.5011			0.00668	$\beta_{w3}=1001.696$			
24956.06	−12448.77	0.531974	24442.76	−12192.13	0.552775	23569.91	−11755.7	0.482901	w_1	REBMIX		
		0.050581			0.055568			0.028813	w_2			
		1.194232			1.171785			1.24847	$\alpha_{w1}=0.5426926$			
		1107.922			1210.845			960.134	$\beta_{w1}=86.28743$			
		2.285546			0.65875			0.538084	$\alpha_{w2}=0.7855334$			
		10598.19			6943.286			3934.514	$\beta_{w2}=609.64774$			
		1.703863			1.703863			1.703863	$\alpha_{w2}=0.5316595$			
		755.6009			796.444			729.4766	$\beta_{w2}=1001.696$			
$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.1, w_2=0.5, w_3=0.3, w_4=0.1$					4	
22217.665	−11068.61	0.309723	22223.125	−11071.34	0.283224	22299.714	−11109.635	0.294332	w_1	EM		
		0.307399			0.319188			0.289847	w_2			
		0.205029			0.195475			0.343583	w_3			
		0.4107735			0.397166			0.44184	$\alpha_{w1}=0.511479$			
		104.81245			124.6144			93.84846	$\beta_{w1}=156.2808$			
		0.6650597			0.57044			0.889702	$\alpha_{w2}=0.77344$			
		413.21926			458.4045			344.29452	$\beta_{w2}=559.484$			
		0.7012848			0.612386			0.802446	$\alpha_{w3}=0.52639$			
		1298.6033			793.2208			1334.5514	$\beta_{w3}=656.8403$			
		0.5094112			0.522664			0.5201218	$\alpha_{w4}=0.62661$			
		1989.6697			2011.473			298.48	$\beta_{w4}=1683.936$			

الخليط	K	خوارزمية التقدير	المعلمات الفعلية	التقديرات (أوزان متفاوته)	LL	BIC	التقديرات (أوزان متساوية)	LL	BIC	التقديرات (أوزان متقاربة)	LL	BIC
		REBMIX	w_1	0.5399099	−12189.39	24459.22	0.547134	−12521.05	25122.54	0.947571	−11966.08	24012.61
			w_2	0.0587055			0.053782			0.007136		
			w_3	0.0248467			0.036418			0.005617		
			$\alpha_{w1} = 0.511479$	1.248434			1.220132			0.942812		
			$\beta_{w1} = 156.2808$	931.9328			1011.734			1500.595		
			$\alpha_{w2} = 0.77344$	1.978265			2.207048			0.605169		
			$\beta_{w2} = 559.484$	7052.628			10436.58			17935.96		
			$\alpha_{w3} = 0.52639$	8.891712			5.541217			13.86289		
			$\beta_{w3} = 656.8403$	42375			43437.17			42549.09		
			$\alpha_{w4} = 0.62661$	1.703863			1.703863			2.037234		
			$\beta_{w4} = 1683.936$	538.3963			597.8005			1272.758		

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

الخليط	k	خوارزمية التقدير	المعلمات الفعلية	التقديرات (أوزان متفاوته)	LL	BIC	التقديرات (أوزان متساوية)	LL	BIC	التقديرات (أوزان متقاربة)	LL	BIC
Weibull	2			$w_1 = 0.8, w_2 = 0.2$			$w_1 = w_2 = 0.5$			$w_1 = 0.6, w_2 = 0.4$		
		EM	w_1	0.73356	−18413.562	36866.245	0.41769	−19284.502	38608.123	0.60488	−18999.003	38037.127
			$\alpha_{w1} = 0.72738$	0.707042			0.73807			0.70009		
			$\beta_{w1} = 409.1349$	391.3087			373.6375			418.264		
			$\alpha_{w2} = 0.60222$	0.530888			0.538066			0.52446		
			$\beta_{w2} = 1335.6705$	939.5847			1061.087			1109.089		
		REBMIX	w_1	0.92415	−18976.1	37991.33	0.94019	−20148.31	40335.74	0.963689	−20273.85	40586.82
			$\alpha_{w1} = 0.72738$	1.14575			1.08331			1.11867		
			$\beta_{w1} = 409.1349$	550.014			1109.504			1157.048		
			$\alpha_{w2} = 0.60222$	1.15285			0.834924			2.91056		
			$\beta_{w2} = 1335.6705$	5536.83			7889.52			10950.53		
	3			$w_1 = 0.6, w_2 = 0.3, w_3 = 0.1$			$w_1 = w_2 = w_3 = 0.333$			$w_1 = 0.4, w_2 = 0.35, w_3 = 0.25$		
		EM	w_1	0.349702	−15912.017	31886.626	0.415011	−17535.254	35133.101	0.29773	−17153.748	34370.089
			w_2	0.457871			0.513857			0.54011		
			$\alpha_{w1} = 0.5426926$	0.493093			0.421895			0.446419		
			$\beta_{w1} = 86.28743$	27.97724			357.3099			120.1621		
			$\alpha_{w2} = 0.7855334$	0.734631			0.563487			0.563188		
			$\beta_{w2} = 609.64774$	448.06445			360.04999			438.5974		
			$\alpha_{w3} = 0.5316595$	0.3848103			0.3154255			0.349781		
			$\beta_{w3} = 1001.696$	310.32132			43.60143			470.8009		
		REBMIX	w_1	0.602934	−18403.86	36870.32	0.447978	−21243.54	42549.67	0.436322	−21037.87	42138.33
			w_2	0.063404			0.026043			0.024578		
			$\alpha_{w1} = 0.5426926$	1.181035			1.220685			1.215898		
			$\beta_{w1} = 86.28743$	511.4924			1920.93			1796.534		
			$\alpha_{w2} = 0.7855334$	0.755605			0.5818297			0.561619		

BIC	LL	التقديرات (اوزان متقاربة)	BIC	LL	التقديرات (اوزان متساوية)	BIC	LL	التقديرات (اوزان متفاوته)	المعلمات الفعلية	خوارزمية التقدير	k	الخليط
		4927.602			5489.268			2959.459	$\beta_{w2}=609.64774$			
		1.703863			1.703863			1.703863	$\alpha_{w2}=0.5316595$			
		1450.49			1529.321			319.4476	$\beta_{w2}=1001.696$			
$w_1=0.2, w_2=0.3, w_3=0.3, w_4=0.2$			$w_1=w_2=w_3=w_4=0.25$			$w_1=0.1, w_2=0.5, w_3=0.3, w_4=0.1$					4	
36851.373	−18382.654	0.352297	36886.001	−18399.973	0.214219	37073.355	−18493.645	0.23866	w_1	EM		
		0.339647			0.569459			0.274197	w_2			
		0.2563501			0.136544			0.34328	w_3			
		0.4493912			0.483448			0.822484	$\alpha_{w1}=0.511479$			
		114.19554			22.75509			202.303	$\beta_{w1}=156.2808$			
		0.7516037			0.773137			0.890822	$\alpha_{w2}=0.77344$			
		434.28058			486.0808			1000.504	$\beta_{w2}=559.484$			
		0.7601273			1.124921			0.481915	$\alpha_{w3}=0.52639$			
		2114.9923			4010.5766			500.4805	$\beta_{w3}=656.8403$			
		0.3768194			0.4337833			0.380324	$\alpha_{w4}=0.62661$			
		976.247			145.9927			506.3257	$\beta_{w4}=1683.936$			
42862.41	−21388.17	0.4316452	42861.46	−21387.7	0.432652	42302.88	−21108.41	0.400338	w_1	REBMIX		
		0.0273378			0.034132			0.019233	w_2			
		0.0170776			0.022812			0.013866	w_3			
		1.234054			1.231689			1.260678	$\alpha_{w1}=0.511479$			
		1672.359			1570.539			1702.252	$\beta_{w1}=156.2808$			
		2.044558			1.71645			1.766723	$\alpha_{w2}=0.77344$			
		12705.5			12680.79			13900.52	$\beta_{w2}=559.484$			
		6.433462			4.879662			7.881071	$\alpha_{w3}=0.52639$			
		129222.4			123588.5			127955.2	$\beta_{w3}=656.8403$			
		1.703863			1.703863			1.703863	$\alpha_{w4}=0.62661$			
		984.4584			920.3967			984.4584	$\beta_{w4}=1683.936$			

6. التعليق على نتائج المحاكاه:

ان الهدف الرئيسى للدراسة هو بناء مجموعة من التوزيعات المختلطة لنمذجة حجم الخسارة المتوقعة والتوصل الى افضل طريقة تقدير ممكنة من بين الطريقتين المقترحتين وقد توصلت المحاكاه الى الآتى:

1- المعلمات المستخرجة باستخدام خوارزمية REBMIX تعطى نتائج أفضل من خوارزمية EM فيما يتعلق بجودة التوافيق ، يظهر ذلك من خلال ملاحظة الزيادة فى قيمة كل من لوغاريتم الامكان LL ومعيار المعلومات البيزى BIC.

2- كلما كانت الأوزان النسبية لمكونات الخليط متساوية أو متقاربة كلما كانت النتائج أقرب الى المعلمة الحقيقية ، والعكس صحيح ، فكلما تفاوتت أوزان الخليط كلما انخفضت معها دقة التقدير يتضح ذلك من خلال كل من قيم LL و BIC.

3- كلما ازدادت أحجام العينات ازداد معها قيم كل من لوغاريتم الامكان LL ومعيار المعلومات البيزى BIC ايضا يظهر الارتفاع الشديد فى النتائج لخوارزمية REBMIX وخوارزمية EM حيث يلاحظ زيادة كبيرة فى قيمتى معيارى المقارنة.

4- بنفس الطريقة ، فإنه بزيادة عدد المكونات تزداد قيم كل من لوغاريتم الامكان LL ومعيار المعلومات البيزى BIC.

5- أثناء عملية المعالجة باستخدام حزم برنامج R i386 4.0.3 لوحظ أن التقدير باستخدام خوارزمية REBMIX يعطى نتائج اسرع بكثير من استخدام خوارزمية EM.

6- وفقا لنتائج المحاكاه فإن افضل مزيج للبيانات محل الدراسة يعبر عن توزيع شدة أو حجم الخسارة هو توزيع وايبل بمكونين ، ويمكن الاستدلال على ذلك من قرب نتائج التقدير من المعلمات الحقيقية المفترضة.

المراجع:

https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture6.pdf. (2020, October 14).

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, Free University of Berlin.

Ghojogh, B., & et al. (2019, jan 20). Fitting A Mixture Distribution to Data: Tutorial.

Hesse, C., Holtackers, D., & Heskes, T. (December 7-8, 2006). On the use of mixtures of Gaussians and mixtures of generalized exponentials for modelling and classification of biomedical signals. *Belgian Day on Biomedical Engineering IEEE Benelux EMBS Symposium.*

Lai, C.-D., & Xie, M. (2006). *Stochastic Ageing and dependence for reliability.* New York, United States: Springer+Business Media Inc.

Nagode, M. (2013). rebmix: An R Package for Continuous and Discrete Finite Mixture Models.

Nagode, M. (2015, April). Finite Mixture Modeling via REBMIX. *Journal of Algorithms and Optimization, 3*(2).

Nagode, M., & Fajdiga, M. (2006, April). An alternative perspective on the mixture estimation problem. *Reliability Engineering & System Safety, 91*(4).

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, Boca Raton.

Panić, B., & at el. (2020, March). Improved Initialization of the EM Algorithm for Mixture Model Parameter Estimation. *Mathematics, 8*(3).

Volodymyr, M., & Ranjan, M. (2010). Finite mixture models and model-based clustering. *Statistics Surveys, 4*.

Zhang, H., & Huang, Y. (2015, March 26). Finite Mixture Models and Their Applications: A Review. *Austin Biom and Biostat, 2*(1).