بعض المقدرات الحصينة والكفوءه لمقياس المنوال في حالة البيانات المستمرة

د.حمزة اسماعيل شاهين [*]　　　　د.رعد فاضل حسن [**]

المستخلص

على الرغم من أن أكثر مقاييس النزعة المركزية شيوعاً للبيانات المستمرة هو مقياس المنوال (*Mode*) إلا أن مقياسي المتوسط (*Mean*) والوسيط (*Median*) لازال هما الأكثر شيوعا لوصف قيمة مقياس الموقع أو المعدل (*Average*) ويعود ذلك لبساطة هذين المقياسين والى إمكانية تقديرهما . وان الوسيط غالباً ما يستخدم بدلاً من المتوسط في حالة تماثل البيانات ويعود سبب ذلك إلى أن الوسيط يكون قريب من المنوال وانه غير حساس للقيم المتطرفة في العينة.

إن هذا البحث يقترح مقدرين للمنوال للبيانات الأصلية , المقدر الأول للمنوال يعتمد على تقدير المتوسط والانحراف المعياري بالاعتماد على المتوسط والانحراف المعياري للعينة , آما المقدر الثاني والذي يتصف بأنه أكثر حصانة فيقوم على تقدير المتوسط بالوسيط والانحراف المعياري بدلالة الانحراف المعياري المطلق للوسيط

ABSTRACT

Although a natural measure of central tendency of a sample of continuous dat its mode , the mean and the median are the most popular measures of locat due to there simplicity and ease of estimation , the median is often used inst of mean for symmetric data because its closer to the mode and its insensitiv extreme values in the sample , Tow estimators of the mode of the original d are proposed , a simple estimator based on estimating the mean by the sam mean and standard deviation by the sample standard deviation , and the ot more robust estimator based on estimating the mean by the median and standard deviation by the standardized median absolute deviation.

* استاذ مساعد/قسم الاحصاء/كلية الادارة والاقتصاد/الجامعة المستنصرية

** استاذ مساعد/كلية المأمون الجامعة/قسم الحاسبات

المقدمة :

على الرغم من أن الكثير من مقاييس الموقع (Location) قد تم إيجادها في السنوات الأخيرة إلا انه لازال الكثير من الباحثين مستمرين على استخدام المتوسط (Mean) والوسيط (Median) لوصف قيمة مقياس الموقع أو المعدل (Average) للبيانات المستمرة (Continuous Data) وبشكل كبير جداً ويعود سبب ذلك إلى أن هذين المقياسين يتصفان بسهولة الفهم وإمكانية التقدير. وان فكرة المنوال (Mode) هي أيضا واضحة ومفهومة حيث يعتبر المنوال أكثر جاذبية كقيمة مناسبة على الأكثر . ولكن طرق تقدير المنوال في حالة البيانات المستمرة ليست واسعة الانتشار حيث أن الكثير من الباحثين يصف المعدل للبيانات المستمرة باستخدام المتوسط ماعدا الحالات التي تكون فيها البيانات ذات (التواء . تفلطح) عالي أو ملوثة بقيم شاذة فانه يكون من المناسب استخدام الوسيط على الأغلب في مثل هذه الحالات. ويعتبر الوسيط اختيار جيد في الحالات المشار إليها أعلاه لسببين. الأول انه في حالات معينة يكون المتوسط غير معول عليه كثيرا وغير معروف والسبب الثاني في حالة تماثل البيانات (Symmetric) فان الوسيط يكون قريب جداً من المتوسط . وعلى الرغم من أن الوسيط يكون دائماً قريب من المنوال إلا انه يفضل في الكثير من الحالات إلى تقدير المنوال حيث يعتبر المنوال من أكثر مقاييس النزعة المركزية شيوعاً وهو يمثل أكثر قيمة نظامية للبيانات ومع ذلك فأن المقدرات السابقة للمنوال كانت تتصف بتحيز عالي وبكفاءة ضعيفة (تباين عالي) وبحساسية للقيم الشاذة وان هذه المحددات قد أعطت تفسيراً لمسألة إهمال المنوال لوصف النزعة المركزية . ولغرض التوسع في إمكانية استخدام المنوال فان هذا البحث يهدف إلى تقدير المنوال ولإيجاد مقدرات تتصف بتحيز قليل وبكفاءة وبحصانة عالية للقيم الشاذة وذلك في حالة البيانات المستمرة التي تتصف بحالة التماثل.

1- خطوات تقدير المنوال

أن الخطوات المقترحة لتقدير المنوال يمكن أن تتلخص بالخطوات التالية:-

1-تحويل البيانات وبشكل تقريبي إلى حالة التوزيع الطبيعي

2-تقدير المتوسط والانحراف المعياري للبيانات المحولة

3-وعلى افتراض أن البيانات المحولة مسحوبة من التوزيع الطبيعي فأنه يتم استخدام المتوسط والانحراف المعياري المقدرين للبيانات المحولة في تقدير المنوال للبيانات الأصلية. والتحويلات المستخدمة في هذا البحث هي تحويلات القوى وحسب الصيغة الآتية

$$y(x;\alpha) = x^{\alpha} \quad ....(1)$$

حيث أن:

y : متغير التحويل

x : المتغير الأصلي

α : عبارة عن ثابت حقيقي لا يساوي صفرا

وهذا يتطلب أن تكون قيم المتغير الأصلي اكبر من الصفر أي أن $(x > 0)$ ، ويمكن تعميم التحويل (1) باستخدام الصيغة التالية

$$y(x;\alpha;\beta) = (x+\beta)^{\alpha} \quad ....(2)$$

حيث أن التحويل (2) سوف يسمح بوجود قيم سالبة للمتغير (x) هذا وان البيانات الأصلية سوف يعبر عنها بالمجموعة $\{x\}_{i=1}^{n}$

أما مجموعة البيانات المحولة فيعبر عنها بالمجموعة $\{y_i(\alpha)\}_{i=1}^{n}$ حيث أن : $y_i(\alpha) = x_i^{\alpha}$.

وان قيمة (α) يتم اختيارها بشكل يجعل البيانات المحولة قريبة جداً من حالة التوزيع الطبيعي ، وبالرغم أن المتغير (y) لايتبع التوزيع الطبيعي بشكل دقيق إلا أنه يمكن وصفه طبيعي وبشكل تقريبي من خلال الاعتماد على اختيار قيمة (α) . وبالتالي فأنه لغرض أجراء التقدير سوف يفترض انه يتبع التوزيع الطبيعي .

فإذا كان المتغير (y) يتبع التوزيع الطبيعي بالمعلمتين $(\bar{y},\delta)$ فان دالة الكثافة الاحتمالية (PDF) للمتغير (y) تكون:-

$$f_y(y;\bar{y};\delta) = \left(\sqrt{2\pi}\delta\right)^{-1}\exp\left[-\frac{(y-\bar{y})^2}{2\delta^2}\right] \quad ....(3)$$

وبالتالي فان دالة الكثافة الاحتمالية (PDF) للمتغير (x) تكون

$$f_x(x;\bar{y};\delta;\alpha) = f_y(x^{\alpha};\bar{y};\delta)\left|\frac{\partial y}{\partial x}\right| = (\sqrt{2\pi}\delta)^{-1}|\alpha|x^{\alpha-1}\exp\left[-\frac{(x^{\alpha}-\bar{y})^2}{2\delta^2}\right]$$

......(4)

ومن المعادلة (4) يمكن وبشكل تقريبي الحصول على عدة توزيعات حيث (α) تعبر عن الالتواء (Skewness) فإذا كانت $\alpha = 1$, فأن للالتواء صفر والتوزيع متماثل . وان المنوال للمتغير

(x) والذي يعبر عنه بالرمز (M) هو عبارة عن قيمة (x) التي تجعل دالة الكثافة الاحتمالية (PDF) أعظم ما يمكن .

وهذا يتطلب اشتقاق دالة الكثافة الاحتمالية بالنسبة للمتغير (x) ومساواتها للصفر أي أن

$$\left[\partial f_{x}(x;\bar{y};\delta;\alpha)/\partial x\right]=0$$

حيث نحصل:-

$$M=\left[\frac{1}{2}\left(\bar{y}+\sqrt{\bar{y}^{2}+\frac{4\delta^{2}(\alpha-1)}{\alpha}}\right)\right]^{1/\alpha} \quad \text{..............(5)}$$

((مع ملاحظة انه في حالة $\alpha = 1$ فمن المعروف انه في حالة التوزيع الطبيعي فان المنوال يكون مساوي إلى المتوسط $\bar{y}$))

وحسب المعادلة (5) يمكن تقدير المنوال وذلك باستبدال المعلمتين $\delta, \bar{y}$ بتقديرهما من بيانات العينة المحولة $\{y_i(\alpha)\}_{i=1}^{n}$.

هذا وعندما تكون قيمة α صغيرة وموجبة فان قيمة الجذر التربيعي في المعادلة (5) تصبح سالبة وهذا الشيء يحدث في حالة العينات الصغيرة ذات الالتواء العالي عندئذ فان تقدير المنوال يكون اقل قيمة في العينة . وان هذه الطريقة سوف يتم توضيحها بشكل مفصل في الفقرة اللاحقة من هذا البحث عند تقدير المنوال باستخدام عدد من الطرق الإحصائية.

3- مقدرات المنوال Estimators of the mode

3-1المقدر ألمعلمي القياسي Standard parametric estimator

أن تقنية تقدير المنوال المذكورة في الفقرة (2) سابقا تكون حسب الخوارزمية التالية :-

1-تحويا البيانات باستخدام قيمة α التي تعظم معامل الارتباط القياسي مابين البيانات المحولة المرتبة والاحصائات المرتبة المتوقعة للتوزيع الطبيعي .

2- تقدير كل من المتوسط والانحراف المعياري للبيانات المحولة باستخدام المتوسط والانحراف المعياري للعينة

3-وبعد التعويض عن $\delta, \bar{y}$ بمتوسط العينة والانحراف المعياري للبيانات المحولة في المعادلة (5) نحصل على تقدير المنوال للبيانات الأصلية

أن الخطوة الأولى من الخوارزمية تتطلب حساب معامل الارتباط لبيرسون من مجموعة البيانات $\{y_i(\alpha)\}_{i=1}^{n}$ المرتبة كالآتي .

$y_1(\alpha) \le y_2(\alpha) .... \le y_n(\alpha)$ ومجموعة البيانات $\{z_i\}_{i=1}^{n}$ التي تمثل الاحصائات المرتبة المتوقعة وفق دالة الكثافة التجميعية (CDF) (Φ) لدالة التوزيع الطبيعي القياسي حيث:

$$z_i = \phi^{-1}\left(\frac{i-(1/2)}{n-1}\right) \qquad ....(6)$$

وان معامل الارتباط يمكن توضيحه كالآتي :-

$$r_{(\alpha)} = \frac{S_+^2(\alpha) - S_-^2(\alpha)}{S_+^2(\alpha)\,\mu\, S_-^2(\alpha)} \qquad ....(7)$$

حيث أن:-

$$S\pm(\alpha) = \delta\left(\frac{y_i(\alpha)}{\delta y_i(\alpha)} \pm \frac{z_i}{\delta z_i}\right) \qquad ....(8)$$

وان العملية δ تمثل الانحراف المعياري للعينة وعلى سبيل المثال :

$$\delta y_i(\alpha) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[y_i(\alpha) - \frac{1}{n}\sum_{i=1}^{n} y_i(\alpha)\right]^2} \qquad ....(9)$$

وعلى افتراض أن α0 هي قيمة α التي تجعل r(α) أعظم ما يمكن وبالتالي تكون هناك قيمة عظمى واحدة فقط للبيانات التي تتبع أنموذج التوزيع المنفرد حيث r(α) تكون رتيبة متناقصة عندما تصبح البيانات المحولة طبيعية بشكل ضعيف . وطريقة حساب α0 هي كما موضحة في الملحق .

أن التحويل $y_i(\alpha) = x_i^{\alpha_0}$ يضمن أن البيانات المحولة تكون قريبة جداً من التوزيع الطبيعي وبالتالي فان المتوسط والانحراف المعياري للعينة $\{y_i(\alpha_0)\}_{i=1}^{n}$ يمكن أن تستخدم في المعادلة (5) لتقدير المنوال والذي يمثل المنوال لتوزيع البيانات للعينة الأصلية $\{x_i\}_{i=1}^{n}$

إن هذا المقدر للمنوال يطلق عليه المنوال ألمعلمي القياسي ويرمز له (SPM) ويتصف ببساطته وكفاءته في حالة كون البيانات $\{y_i(\alpha_0)\}_{i=1}^{n}$ طبيعية بشكل تقريبي . ومن جانب أخر فان هذا المقدر يكون غير حصين (Robust) للقيم الشاذة حيث في حالة وجود قيمة كبيرة من قيم (x) عندئذ فان قيمة α0 تتجاوز أي حدود موضوعة لها وبالتالي فان التقدير يكون عديم الفائدة ولا يخدم الغرض الذي وجد من اجله .

3-2 المقدر المعلمي الحصين Robust parametric estimator

أن الخطوات الواردة في الخطوة (2) من هذا البحث لحساب المنوال تكون ذات حصانة عالية في حالات تلويث البيانات أذا أخذت بنظر الاعتبار النقاط التالية :-

1-أن يتم تحويل البيانات باستخدام قيمة α التي تجعل معامل الارتباط الحصين مابين بيانات التحويل المرتبة والاحصائات المرتبة المتوقعة للتوزيع الطبيعي أعظم ما يمكن .

2-إن يتم تقدير المتوسط والانحراف المعياري للبيانات المحولة باستخدام الوسيط والانحراف المطلق المعياري للوسيط (MAD)

3-أن يتم التعويض عن $\delta, \bar{y}$ في المعادلة (5) بكل من الوسيط والانحراف المطلق المعياري للوسيط (MAD) للبيانات المحولة لغرض تقدير المنوال للبيانات الأصلية .

وان معامل الارتباط الحصين يعتمد على معامل الارتباط الخطي العام وكذلك على معلمة القياس δ في المعادلة (8) التي نمثل الانحراف المعياري وبإعادة كتابة $y_1(\alpha) \le y_2(\alpha) \le .... \le y_n(\alpha)$ فان معامل الارتباط المستخدم يكون كالأتي :-

$$R(\alpha) = \frac{S_+^2(\alpha) - S_-^2(\alpha)}{S_+^2(\alpha) + S_-^2(\alpha)} \qquad ....(10)$$

حيث أن :-

$$S_\pm(\alpha) = \Delta\left(\frac{y_i(\alpha)}{\Delta y_i(\alpha)} \pm \frac{z_i}{\Delta z_i}\right) \qquad ....(11)$$

وان العملية Δ نحصل منها على (MAD) حيث أن $\Delta y = \delta$ إذا كان (y) لا يتبع التوزيع الطبيعي بانحراف معياري δ

وعلى سبيل المثال :-

$$\Delta y_i(\alpha) = \left[1/\phi^{-1}(3/4)\right] med\left|y_i(\alpha) - med(y_i(\alpha))\right|$$
$$= 1.4826 \ \ med\left|y_i(\alpha) - med9y_i(\alpha)\right| \qquad ....(12)$$

حيث med يمثل عملية الوسيط إما $med \ \ y_i(\alpha)$ فتمثل الوسيط لمجموعة البيانات المحولة $\{y_i(\alpha)\}_{i=1}^{n}$ هذا وان قيمة α0 يمكن إيجادها بتعظيم R(α) أي $R(\alpha_0) = \max_{\alpha} R(\alpha)$ وبالنسبة لحالة التوزيع الطبيعي فان المتوسط يكون مساوي للوسيط والانحراف المعياري مساوي إلى الانحراف المطلق المعياري للوسيط (MAD)

وبذلك يتم التعويض عن $\delta, \bar{y}$ في المعادلة (5) بـ ـــــــ $\Delta y_i(\alpha_0), med \ \ y_i(\alpha_0)$ لتقدير المنوال . إن حصانة هذا المقدر للمنوال والذي يطلق عليه بالمنوال المعلمي الحصين (RPM) يمكن قياسها بنقطة الانهيار (Break down point) والتي هي اصغر نسبة مئوية من البيانات الملوثة تجعل المقدر يتخطى الحد المعقول ((ينهار)) إن أفضل نقطة انهيار يمكن توقعها وينشدها الباحثون هي (50%) يصبح من المستحيل التمييز بين الجزء الجيد والجزء غير الجيد (الشاذ) من العينة .

فعلى سبيل المثال في حالة لدينا عينة بحجم (n) من المشاهدات فان نقطة الانهيار لمقدر الوسيط تساوي $(\frac{n+1}{2n})$ في حالة عدد المشاهدات فردي أو تساوي (50%) في حالة عدد المشاهدات الزوجي واعتماداً على الوسيط فأن مقدر المنوال سوف يمتلك نفس نقطة الانهيار والتي تمثل اعلى نقطة انهيار ممكنة لمقياس الموقع وان مقدر المنوال الموضح في الفقرة (3-1) من هذا البحث يعتبر اقل حصانة بسبب أن نقطة انهيار كل من المتوسط والانحراف قد تصل إلى اقل من (1/n) أو أكثر وبالتالي فان وجود قيمة شاذة واحدة في البيانات يجعل قيم المتوسط والانحراف المعياري كبيرة جداً .

3-3 مقدرات Grenander's

هنالك صنف من المقدرات اللامعلمية للمنوال تعرف بمقدرات عائلة Grenander's للمنوال للبيانات $\{x_i\}_{i=1}^{n}$ حيث أن $x_1 \le x_2 \le \dots\dots\dots \le x_n$ وتعطى حسب الصيغة التالية

$$M_{p,k}^{*} = \frac{(1/2)\sum_{i=2}^{n-k}(x_{i+k}+x_i)/(x_{i+k}-x_i)}{\sum_{i=1}^{n-k}1/(x_{i+k}-x_i)^{p}}, 1 < p < k \qquad \dots\dots.(13)$$

حيث أن p,k أعداد حقيقية وتكون ثابتة لكل مقدر .

وان $M_{p,k}^{*}$ يملك نقطة انهيار تصل إلى (k+1)/n وهذه تقترب من الصفر عندما $n \to \infty$ ويتصف $M_{p,k}^{*}$ بنفس خصائص المقدر في الفقرة (3-1) حيث يكون غير حصين للقيم الشاذة

3-4 مقدرات الكثافة اللامعلمية Nonparametric density estimators

المقدرات اللامعلمية المشار إليها في الفقرة (3-3) تعتبر مقدرات مباشرة ولكن هناك مقدرات لامعلمية للمنوال تعتمد على تقدير دالة الكثافة الاحتمالية (PDF) حيث يمكن تقدير المنوال من خلال تعظيم دالة الكثافة التجريبية التمهيدية (EDF) وبالنتيجة يتم تقدير (PDF) وان دالة الكثافة التجريبية التمهيدية (EDF) بالاعتماد على الدالة (Normal kernel) هي كالآتي

$$\hat{f}(x) = \frac{1}{nh}\sum \exp\left[-\frac{1}{2}\left(\frac{(x-x_i)^2}{h}\right)\right] \qquad \dots\dots.(14)$$

وعند اخذ قيم صغيرة لمعلمة التمهيد (h) نحصل على تحيز قليل ولكن بتباين عالي عند تقدير المنوال وان تقدير المنوال يتم باستخدام دالة الكثافة التجريبية (EDF) حيث يرمز للمنوال بالرمز $\overline{M}$ ويعرف كالآتي :-

$$\hat{f}(\overline{M}) = \max_x \hat{f}(x) \qquad \text{........(15)}$$

4-المحاكاة Simulation

لقد تم استخدام المحاكاة لغرض تقدير المنوال بالاعتماد على الطرق الموضحة في الفقرة (3) من هذا البحث حيث تم استخدام برنامج مكتوب بلغة باسكال في توليد عينات تعود إلى التوزيع الطبيعي (Normal dist) والتوزيع ألوغاريتمي (Lognormal dist)

وتوزيع باريتو (Pareto dist) الذي كانت دالة الكثافة الاحتمالية له كالأتي

$$f(x) = \begin{cases} 1/2x^{3/2} & if \quad x \geq 1 \\ 0 & if \quad x < 1 \end{cases}$$

وقد تم افتراض مجموعة من قيم معالم للتوزيعات أعلاه والمدرجة في الجدول أدناه

جدول رقم (1)

المعالم الافتراضية للتوزيعات (الطبيعي , ألوغاريتمي , باريتو)

المعالم / التوزيع	M	δ	Me	M0
الطبيعي	3	1	3	3
ألوغاريتمي	2	1	2.72	2
باريتو	N/A	N/A	4	1

حيث تعذر افتراض قيم للمتوسط والانحراف المعياري لتوزيع باريتو حيث انه معروف أن متوسط توزيع باريتو يكون غير محدد (infinite) .

ولكل توزيع تكررت التجربة (100) مرة وبأحجام عينات مختلفة n = 30 , 50 , 100 ولكل عينة من هذه العينات تم تقدير المنوال باستخدام المقدرات الموضحة في الفقرات (1-3) , (2-3) , (3-3) , (4-3) , ولغرض إجراء المقارنة تم تقدير المتوسط والوسيط للعينات المولدة وتم حساب التحيز (Bias) الذي يمثل الفرق بين متوسط التقدير وقيمة المعلمة المقدرة وتم حساب التباين لجميع المقدرات وقد عرضت النتائج في الجداول (2 , 3 , 4) اعتماداً على البيانات المولدة والخالية من التلويث .

جدول رقم (2)

التحيز والتباين لستة مقدرات لمقياس المنوال للعينات المولدة بحجم (20) مشاهدة والمسحوبة من ثلاث توزيعات مستمرة

المعالم المقدرة	التوزيعات		
	Normal dist	Lognormal dist	Pareto dist
Mean	-0.00213 (0.0126)	0.0911 (3.214)	N/A N/A
Median	-0.00201 (0.09432)	0.0132 (0.6413)	0.2621 (7.421)
Standard parametric mode (SPM)	-0.03901 (0.1131)	0.1141 (0.1262)	0.1341 (0.0251)
Robust Parametric mode (RPM)	-0.012 (0.2210)	0.3361 (0.1911)	0.1101 (0.003)
Grenander $M^*_{p,k}$	-0.0178 (0.3321)	1.024 (0.3411)	1.426 (2.536)
Empirical density model	-0.6213 (0.1921)	1.001 (0.2130)	1.3242 (0.1710)

جدول رقم (3)

التحيز والتباين لستة مقدرات لمقياس المنوال للعينات المولدة بحجم (50) مشاهدة والمسحوبة من ثلاث توزيعات مستمرة

المعالم المقدرة	التوزيعات		
	Normal dist	Lognormal dist	Pareto dist
Mean	-0.0013 (0.031)	0.0782 (4.214)	N/A N/A
Median	-0.0019 (0.0812)	0.2132 (0.423)	0.2132 (6.5524)
Standard parametric mode (SPM)	-0.03061 (0.0061)	0.1211 (0.0425)	0.1399 (0.0112)
Robust Parametric mode (RPM)	-0.013 (0.6234)	0.2214 (0.5211)	0.1065 (0.0621)
Grenander $M^*_{p,k}$	-0.0119 (0.6234)	1.0025 (0.6742)	1.0271 (1.2111)
Empirical density model	-0.5211 (0.1101)	1.001 (0.2770)	1.1259 (0.0332)

جدول رقم (4)

التحيز والتباين لستة مقدرات لمقياس المنوال للعينات المولدة بحجم (100) مشاهدة والمسحوبة من ثلاث توزيعات مستمرة

المعالم المقدرة	التوزيعات		
	Normal dist	Lognormal dist	Pareto dist
Mean	-0.0075952 (0.0117255)	0.00776605 (0.356727)	N/A N/A
Median	-0.0038602 (0.0156868)	0.0359824 (0.123129)	0.223126 (0.674425)
Standard parametric mode (SPM)	-0.00573208 (0.0260692)	0.0604912 (0.373942)	0.460835 (0.01152515)
Robust Parametric mode (RPM)	-0.044038 (0.0647818)	0.0609892 (0.0797819)	0.15173 (0.0608125)
Grenander $M^{*}_{p,k}$	-0.0284455 (0.246342)	1.017056 (1.54099)	0.930797 (0.564143)
Empirical density model	-0.0434379 (0.0743907)	0.695424 (0.0595135)	1.12872 (0.0722581)

الاستنتاجات والتوصيات

يلاحظ من خلال الجداول (4 , 3 , 2) إن مقدر (Grenander) للمنوال يمتلك اكبر تحيز وتباين من بين المقدرات الستة للمنوال ولجميع التوزيعات الثلاث المستخدمة في البحث وان المقدر ألمعلمي القياسي للمنوال (SPM) يعتبر هو الأفضل من بين المقدرات الأخرى للمنوال ما عدا المقدر ألمعلمي الحصين (RPM) حيث يمتلك اقل تباين وتحيز بالنسبة لتوزيع (Pareto) عندما يكون حجم العينة (20) وبالاعتماد على هذه النتائج فان المقدر ألمعلمي القياسي (SPM) يعتبر هو أفضل اختيار لمقدر المنوال في حالة البيانات المستمرة غير الملوثة ما عدا الحالات التي يكون التوزيع ذو التواء عالي أو ذا ذيول طويلة فان لمثل هذه الحالات يكون المقدر ألمعلمي الحصين للمنوال (RPM) هو الأفضل .

ونوصي بتطبيق مقدرات المنوال المقترحة في هذا البحث على البيانات الملوثة التي تتضمن القيم الشاذة .

3-Granander, U.(1965),"Some direct estimates of the mode" *Ann.math.statist*36,131-138

4-Huber ,P.J,(1981), *Robust Statistics* ,John Wiley & Sons (New York)

5-Rousseeuw,P.J and Croux, C ,(1993)'Alternative to the median absolute Deviation " *,J,Amm,Statist.Assoc* 88,1273-1283 .

الملحق

خوارزمية إيجاد أس التحويل α_0

بافتراض أن الدالة $\lambda(\alpha)$ التي هي بمثابة كل من الدالتين $r(\alpha)$ أو $R(\alpha)$ تكون أعظم مايمكن عند نقطة وحيدة $(\alpha 0)$ وتتصف بأنها متزايدة على وتيرة واحدة عند $\alpha \le \alpha_0$ ومتناقصة بوتيرة واحدة عند $\alpha \ge \alpha_0$.

ولحساب قيمة $(\alpha 0)$ فانه يجب أولا إيجاد ثلاث قيم $\alpha_1, \alpha_2, \alpha_3$ بحيث تحقق

$\alpha_1 < \alpha_2 < \alpha_3$ و $\lambda(\alpha_1) < \lambda(\alpha_2)$

و $\lambda(\alpha_3) < \lambda(\alpha_2)$ وهذا يضمن تحقق $\alpha_1 < \alpha_0 < \alpha_3$

وفي المحاكاة التي أجريت في هذا البحث تم افتراض $a1 = -1$ ، $a2 = 1$ وان $a3 = 2.1$ وقد استعملت هذه كتخمينات أولية حيث أن تناقص قيمة $a1$ أو تزايد قيمة $a3$ يضمن تحقق

$\lambda(\alpha_1) < \lambda(1)$ و $\lambda(\alpha_3) < \lambda(1)$

وان استعمال قيم غير تكاملية بالنسبة إلى $a3$ هو لغرض تجاوز الصعوبات التحليلية عند حساب قيمة $\lambda(0)$ في الخوارزمية الآتية :

إن مجادلة خوارزمية $\max(\alpha_1, \alpha_3)$ بالرجوع إلى $a0$ سوف يتم من خلال المستوى المرغوب للدقة ((مستوى الدقة المستخدم في البحث هو 0.0001)) .

مجادلة لـــــ $\max(\alpha_1, \alpha_3)$ **Argument for** $\max(\alpha_1, \alpha_3)$

1- إذا كان $A_5 - A_1 \le 0.0001$ فانه يكون بالرجوع إلى $A_1 + A_2/2$ ونتوقف أو لا ننتقل إلى الخطوة (2)

2- يتم تقسيم مجال $[A1,A5]$ إلى أربع فترات متساوية البعد هي $[A1,A2]$ و $[A2,A3]$ و $[A3,A4]$ و $[A4,A5]$ بحيث تحقق الآتي

$$A_2 - A_1 = A_3 - A_2 = A_4 - A_3 = A_5 - A_4$$

3- يتم حساب الفرق di إلى $\lambda(\alpha)$ وللفترات الأربعة حيث

$$d_i = \lambda(A_{i+1}) - \lambda(A_i) \qquad i = 1,2,3,4$$

4- فإذا كان $d_j \geq 0$ و $d_{j+1} \geq 0$ فانه يكون معروفا $\lambda(A_j) < \lambda(\alpha_0)$ وان $\lambda(A_{j+1}) < \lambda(\alpha_0)$ وهذا سوف يحقق مجادلة $\max(A_j, A_{j+2})$ تحقق الشروط المطلوبة بالعودة إلى $a0$ وماعدا ذلك ننتقل إلى الخطوة (5)

5- إذا كان $\lambda(A_1) < \lambda(A_5)$ يكون الرجوع إلى مجادلة $\max(A_4, A_5)$ ماعدا ذلك يكون الرجوع إلى المجادلة $\max(A_1, A_2)$