اقتراح خوارزمية لإيجاد معاملات نموذج الانحدار الخطي الضبابي من خلال دمج خوارزميتان حسابات بايز التقريبية ABC مع معاينة جيبس

لؤي فرح * سعد الدين العبدالله** وفاء عيسى**

(الإيداع: 12 آيار 2022 ، القبول: 24 تموز 2022)

الملخص

تم تطوير مفهوم نموذج الانحدار الخطي لمعالجة أنواع معقدة من الظواهر المدروسة، فقد تم نمذجة هذه الظواهر باستخدام الانحدار الخطي الضبابي بمدخلات قطعية ومخرجات ضبابية، ولتقدير معالم هذا النموذج يوجد لدينا منهجين هما: الإمكانية و المربعات الصغرى.

قمنا في هذا البحث باقتراح خوارزمية تدمج بين منهج الإمكانية (نموذج Tanaka) ومنهج المربعات الصغرى (نموذج Zeng) لإيجاد معاملات الانحدار الخطي الضبابي عبر تطبيق منهجية بايز التي تأخذ عشوائية الظواهر المدروسة بالحسبان وتتميز بقدرة عالية على التنبؤ، وذلك من خلال دمج خوارزمية معاينة Gibbs مع خوارزمية حسابات بايز التقريبية، وافتراض قيم ابتدائية وتوزيعات القبلية.

وتم التحقق من فعالية هذه الخوارزمية من خلال تطبيقها على بيانات واقعية وبيانات مولدة، كما تمت مقارنة النتائج مع نموذج الإمكانية لتاناكا Tanaka possibilistic model ونموذج Zeng باستخدام مقياس جودة التلاءم(GOF).

وقد أعطت الخوارزمية المقترحة أفضل دقة لتقدير معاملات نموذج الانحدار الخطي الضبابي، كما أنها تغلبت على مشكلة تحديد دالة الإمكانية في النموذج الضبابي وفي تحديد التوزيعات القبلية.

الكلمات المفتاحية: الانحدار الخطي الضبابي، خوارزمية الرفض ABC (حسابات بايز التقريبية)، منهجية بايز، ، معاينة جبيس، نموذج Tanaka، نموذج Zeng .

* طالب دراسات عليا (دكتوراه)–قسم الإحصاء الرياضي–كلية العلوم–جامعة حلب

** قسم الإحصاء الرياضي–كلية العلوم–جامعة حلب

Suggesting an algorithm for estimating the parameters of a linear regression model, Bayesian approximation computation ABC with Gibbs sampler algorithm

Louay Farah * Saad Alden Al-Abdullah **, Wafaa Issa

(Received:12 May 2022 ,Accepted:24 July 2022)

Abstract:

The concept of linear regression model has been developed to address complex types of studied phenomena, these phenomena have been modeled using fuzzy linear with crisp inputs and fuzzy outputs, and to estimate the parameters of this model we have two methods: possibilistic and least squares methods.

In this paper, we have proposed an algorithm that combines possibilistic and least squares methods to find the fuzzy linear regression coefficients by applying the Bayesian Method , which notes the random nature of phenomena and has a high predictive ability, by merging the Gibbs sampler algorithm with the Bayesian approximation algorithm, and assuming the initial values and distributions prior.

Its quality was verified by applying it to real and generated data, and the results were compared with Tanaka possibilistic model and least squares methods (Zeng's model) using the Quality of Fit (GOF) scale.

The results of the research indicated that the proposed algorithm gives the best accuracy for estimating the coefficients of the fuzzy linear regression model, and it also overcame the problem of determining the likelihood function in the fuzzy model and in determining the prior distributions.

Keywords: linear regression model, rejection ABC (approximate Bayesian computation) algorithm, Bayesian Method, Gibbs sampler algorithm, Tanaka possibilistic model,. Zeng's model,

***Postgraduate Student (PhD)-Dept. of Mathematical Statistics -Faculty of Science-University of Aleppo**

****Dept. of Mathematical Statistics-Faculty of Science-University of Aleppo**

1-مقدمة: Introduction

يشكل تحليل وتقييم العلاقات بين مجموعة من المتغيرات الهدف الأساسي لمعظم الأبحاث بغرض الوصول إلى نموذج رياضي يصف هذه العلاقات، ووجود أنواع عديدة من البيانات أدى الى بروز نماذج مختلفة لتمثيل هذه البيانات وتأتي في مقدمة هذه النماذج ما يُعْرَف بنماذج الانحدار (Regression Models) وتضم هذه النماذج متغيرات الإخراج (التابعة أو الاستجابة)،(Dependent Variables) والتي يمكن التنبؤ بها بواسطة متغيرات أخرى تعرف بالمتغيرات المدخلات (المستقلة أو التفسيرية) (Independent Variables) أو المتغيرات التفسيرية (Explanatory Variables)،(نور الدين،2017)(إسماعيل،2001).

يعتبر تحليل الانحدار من أهم الأدوات الإحصائية المستخدمة في العديد من المجالات العلمية وتستخدم نظرية الاحتمالات لمعالجة عدم التأكد الموجود في هذه النماذج الذي يعزى إلى العشوائية، حيث تتعامل مع البيانات على أنها أعداد دقيقة على الرغم من أن هناك الكثير من البيانات التي لا يمكن تقديمها كأعداد دقيقة أو لا يمكن تقديمها بدقة مثل البيانات في شكل لغوي أو صفات مثل ارتفاع درجة الحرارة وانخفاض المرونة وارتفاع ضغط الدم وحتى نتائج قياس الدقة للمتغيرات المستمرة ليست أعداد دقيقة من أجل ذلك ظهرت الحاجة لتمديد فكرة الانحدار الخطي الإحصائي وتجهيزه بالعناصر من الضبابية وعليه فإن نظرية المجموعات الضبابية تكون الوسيلة المناسبة والفاعلة في صياغة النماذج الإحصائية التي تعزي عدم الدقة إلى الضبابية، (الصباغ،2005)(الصباغ ،الغنام،2009).

في عام 1982 كان Tanaka أول من اقترح مفهوم انحدار خطي الضبابي من خلال ضبابية معلمات النموذج، وتم تطوير طرائق الانحدار الخطي لتطبق في مجالات متنوعة مثل تصميم المنتج، التمويل والصناعة التحويلية،...

يصنف النموذج الخطي الضبابي إلى ثلاث أصناف بحسب أنواع بيانات الإدخال والإخراج وهي:

1) المدخلات القطعية والمخرجات القطعية (CICO) (Crisp-Inputs and Crisp-Outputs).

2) المدخلات الضبابية والمخرجات الضبابية (FIFO) (Fuzzy-Inputs and Fuzzy-Outputs).

3) مدخلات قطعية ومخرجات ضبابية (CIFO) (Crisp-Inputs and Fuzzy-Outputs) .

(Wang, Reformat,Yao, Zhao, Chen, 2020)،(Viertl, 2011) .

سنركز في هذا البحث على تحليل الانحدار الخطي الضبابي باستخدام الصنف الثالث CIFO .

إن الاستراتيجيتان الأساسيتان المستخدمتان لتقدير نموذج الانحدار الضبابي هما:

أولاً- منهج الإمكانية الذي يعتمد على خاصية التضمين، بالاستناد لها قدم Tanaka طريقة يحاول من خلالها تقليل الضبابية للنموذج الكلي عن طريق تقليل الفروق الاجمالية للمعاملات الضبابية تحت قيود التضمين كل عينة التدريب داخل مجال محدد للبيانات وعامل الانحدار الضبابي كمسألة برمجة خطية (Linear Programming Problem) (LP) وهذه الطريقة حساسة للغاية للقيم المتطرفة.

ثانياً- منهج المربعات الصغرى الضبابية الذي يعتمد على خاصية الاتجاه المركزي فقد اعتمد عليها Diamond في تقديم نموذجه وهي امتداد لطريقة المربعات الصغرى التقليدية تحاول هذه الطريقة تقليل المسافة التربيعية بين البيانات المشاهدة والقيم المقدرة، إن هذه المنهجية لها عيبان: وجود افتراضات الإحصائية وضعف القدرة التنبؤية بسبب الإفراط بالملائمة وتوجد بعض المحاولات تحاول تحقيق التوازن بين الأسلوبين للحصول على أفضل تقدير للانحدار جميع طرائق السابقة تستخدم حساب الأمثلية لإيجاد التقديرات المطلوبة التي لا تأخذ عشوائية ظواهر بالحسبان في عملية التركيب النموذج وما مدى ثقتنا في النتائج التي تم الحصول عليها غير ملحوظ مع استخدام هذه الأساليب الحتمية للتقدير،(Wang, etal, 2020).

قمنا في هذا البحث على عكس كل الأساليب السابقة ببناء النموذج الضبابي باستخدام منهج بايز وقدمنا محاولة للجمع بين خاصية الاتجاه المركزي لمنهج المربعات الصغرى وخاصية التضمين لمنهج الإمكانية مع انحدار بايز الذي يتمتع بمزايا عديدة تدمج المعلومات الأولية مع عينة المشاهدة وكذلك القدرة على تطوير معلوماتنا بناءً على التوزيع البعدي، وتحسين الأداء للعينات الصغيرة ، وتوليد تنبؤات على الرغم من وجود أمثلة عن الاستدلال الضبابي القائم على القواعد المنهج البايزي ، فإنه لا يوجد حل بايزي لتقدير الانحدار الضبابي الخطي باستثناء محاولة لإيجاد الانتشارات باستخدام حسابات بايز التقريبية، وقد قمنا باقتراح خوارزمية جديد تدمج بين معاينة جيبس وحسابات بايز التقريبية وتقديم حلاً عددياً لتقدير المعالم الضبابية لهذا النموذج.

2-الموادُّ وطرائقُ البحث: Materials and Methods

2-1- تعريف المجموعة الضبابية Definition of Fuzzy Sets :

عرف العالم زاده المجموعة الضبابية على أنّها المجموعة الجزئية A من المجموعة الشاملة U الغير خالية المزودة بدالة عضوية $\mu_A(x): U \to [0,1]$ وقيمة الدالة تمثل درجة عضوية x في المجموعة A فكلما اقتربت قيمة $\mu_A(x)$ من الواحد ارتفعت درجة عضوية x في A .

بالتالي يمكننا القول بأن المجموعة الكلاسيكية حالة خاصة من المجموعة الضبابية وذلك عندما تكون دالة العضوية لها القيمتين 1 أو 0 وتدعى بالمجموعة القطعية لتمييزها عن المجموعة الضبابية عند استخدام المنطق الضبابي، ويمكن وصف المجموعة الضبابية وصفاً تاماً بالشكل:(Zadeh,1965)،(Nguyen, Walker, Walke ,2018).

$$A = \{(x, \mu_A(x)); x \epsilon U\} \tag{1}$$

2-2- مستويات α α –Levels :

لتكن $\boldsymbol{A}$ مجموعة ضبابية جزئية من المجموعة U و$\alpha \epsilon [0,1]$ ، فإن مستويات α للمجموعة الضبابية $\boldsymbol{A}$ هي المجموعة القطعية من المجموعة U ويرمز لها بالرمز $[A]^\alpha$ وتعرف بالشكل:

$$[A]^\alpha = \{x \in U;\ \boldsymbol{\mu_A(x)} \geq \alpha\ \}; 0 < \alpha \leq 1 \tag{2}$$

من أجل $\alpha = 0$ فإن $[A]^0$ تعرف بأنها أصغر مجموعة جزئية مغلقة تحتوي المجموعة المولدة

$$supp(A) = \{x \in U;\ \mu_A(x) > 0\}$$

للمجموعة $\boldsymbol{A}$،(بشر،2019)(Barros, Bassanezi, Lodwick,2017) .

2-3- تعريف العدد الضّبابي Definition of Fuzzy Number:

لتكن $\boldsymbol{A}$ مجموعة ضبابية في مجموعة الأعداد الحقيقة $\mathbb{R}$ يقال عن $\boldsymbol{A}$ إنها عدد ضبابي (Fuzzy Number) (FN) إذا كان من أجل أي$\alpha \epsilon [0,1]$ المجموعة $[A]^\alpha$ فترة غير خالية متراصة، (Barros,etal,2017)(Fullér,1995).

2-3-1- العدد الضّبابي LR fuzzy number LR:

ليكن $\boldsymbol{A}$ عدد ضبابي (FN) عندئذً يدعى العدد الضبابي $\boldsymbol{A}$ بالعدد الضّبابي LR إذا كانت له دالة العضوية التالية:

$$\mu_A(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right) ; m-\alpha \leq x \leq m \\ R\left(\dfrac{x-m}{\beta}\right) ;\ m \leq\ x \leq m+\beta \\ \quad 0 \quad ; \qquad O.W \end{cases} \tag{3}$$

من أجل كل $x \in \mathbb{R}$

حيث:

التابعين $L, R: \mathbb{R} \to [0,1]$ متناقصان ومستمران ويحققان $L(1) = R(1) = 0$ و $L(0) = R(0) = 1$ ، ويدعى L التابع الشكلي اليسار ويدعى R التابع الشكلي اليمين لـ$\boldsymbol{A}$ و m عدد حقيقي تدعى المركز لـ$\boldsymbol{A}$.

$\alpha, \beta > 0$ ، تدعيان α الانتشار اليسار و β الانتشار اليمين $\boldsymbol{A}$ ويرمز للعدد الضبابي LR بالرمز: $\boldsymbol{A} = (m, \alpha, \beta)_{LR}$.

وكحالة خاصة من العدد الضبابي LR العدد الضبابي المثلثي Triangular Fuzzy Number (TFN) ويرمز له بالرمز $\boldsymbol{A} = (m, \alpha, \beta)_T$ والدالة العضوية لها الشكل:

$$\mu_A(x) = \begin{cases} \dfrac{x - m + \alpha}{\alpha} \; ; m - \alpha \le x \le m \\ \dfrac{m + \beta - x}{\beta} ; m \le \; x \le m + \beta \\ \quad 0 \quad ; \qquad O.W \end{cases} \tag{4}$$

وإذا كان $\beta = \alpha = e$ في هذه الحالة يدعى العدد المثلثي الضبابي TFN بالعدد المثلثي الضبابي المتناظر ويرمز له بالرمز $\boldsymbol{A} = (m, e)_T$،(Zimmermann,2011)(Wang,etal,2020).

2-3-2- العمليات على الأعداد الضبابية المثلثية:

ليكن لدينا $\boldsymbol{A} = (c_1, l_1, r_1)_T$ و $\boldsymbol{B} = (c_2, l_2, r_2)_T$ عددين ضبابين مثلثيين وكان $\lambda \epsilon \mathbb{R}$ مغاير للصفر عندئذٍ تعرف العمليات الجبرية كما يلي:

1. **ضرب بعدد سلمي:**

$$\lambda A = \begin{cases} (\lambda c_1, \lambda l_1, \lambda r_1)_T & ; \; \lambda > 0 \\ (\lambda c_1, -\lambda l_1, -\lambda r_1)_T & ; \; \lambda < 0 \end{cases} \tag{6}$$

2. **الجمع:**

$$\boldsymbol{A} + \; \boldsymbol{B} = (c_1 + c_2, , l_1 + l_2, r_1 + \; r_2)_T \tag{7}$$

3. **الطرح:**

$$\boldsymbol{A} - \; \boldsymbol{B} = (c_1 - c_2, l_1 + \; r_2, r_1 + l_2 \,)_T \tag{8}$$

، (Wang,etal,2020)(Fullér,1995).

3-3-2- معايير لقياس الاختلاف بين عددين ضبابيين مثلثين:

هناك عدة معاير كمية لقياس الاختلاف بين عددين ضبابين مثلثين يمكن استخدامها لتقيم أداء نموذج الانحدار الضبابي سنورد فيما يلي ثلاث معاير من أهمها وأكثرها استخداماً:

المعيار الأول: المسافة بين عددين ضبابين مثلثين كما قدمه Diamond.

ليكن لدينا $\boldsymbol{A} = (c_1, l_1, r_1)_T$ و $\boldsymbol{B} = (c_2, l_2, r_2)_T$ عددين ضبابين مثلثيين فإن المسافة Diamond بينهما تعرف بالشكل:

$$D(A, B) = (c_1 - l_1 - c_2 + l_2)^2 + (c_1 - c_2)^2 + (c_1 + \; r_1 - c_2 - \; r_2)^2 \tag{9}$$

المعيار الثاني: معيار التشابه بين $\boldsymbol{A} = (c_1, l_1, r_1)_T$ و $\boldsymbol{B} = (c_2, l_2, r_2)_T$ يعرف كما يلي:

$$S(A,B) = 1 - \frac{|c_1 - l_1 - c_2 + l_2| + |c_1 - c_2| + |c_1 + r_1 - c_2 - r_2|}{3 \times (\max(c_1 + r_1, c_2 + r_2) - \min(c_1 - l_1, c_2 - l_2))} \quad (10)$$

المعيار الثالث: درجة ملائمة النموذج الضبابي كما قدمها العالم Tanaka تقاس بواسطة الدليل $\bar{h}$ يمثل الحد الأدنى تضمين بين المخرجات المقدرة والبيانات المشاهدة ويعرف كما يلي:

$$\bar{h} = \begin{cases} \max\limits_{[B]^h \subset [A]^h} h \;, & if\ \exists h \epsilon [0,1], [B]^h \subset [A]^h \\ 0 & \text{ذلك خلاف} \end{cases} \quad (11)$$

حيث: A المخرجات الضبابية، B المخرجات الضبابية المقدرة

و $[A]^h = \{x;\ \boldsymbol{\mu_A}(\boldsymbol{x}) \geq h\ \}, [B]^h = \{x;\ \boldsymbol{\mu_B}(\boldsymbol{x}) \geq h\ \}$

مستويات h للمجموعتين، (Alstolanty,Alnagash,2015)(Wang,etal,2020) (Wang, Zhang, Mei, 2007).

2−4− تحليل الانحدار:

يُعرَّف تحليل الانحدار بمفهومه الحديث بحسب Fox (1997) كما يلي:

"يختص تحليل الانحدار بدراسة اعتماد متغير يعرف بالمتغير التابع على متغير واحد أو أكثر تعرف بالمتغيرات المستقلة أو التفسيرية، وذلك بغرض التنبؤ بالقيم المتوقعة للمتغير التابع بواسطة المتغيرات التفسيرية.

وبشكل مبسط فأن تحليل الانحدار هو مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات التوضيحية في العلاقة وغالبًا ما تسمى العلاقات من هذا النوع بنماذج الانحدار.

ويمكن التعبير عن شكل العلاقة بين المتغيرات المستقلة ومتغير التابع على شكل معادلة رياضية تدعى نموذج الانحدار كما يلي:

$$Y = m(X) + \epsilon \quad (12)$$

بحيث أن:

Y المتغير التابع أو المتغير المفسَّر

X المتغيرات المستقلة أو المتغيرات التوضيحية

ϵ الخطأ العشوائي غير ملحوظ.

$m(\bullet)$ يدعى تابع انحدار مجهول أو منحني الانحدار.

وعندما يتألف نموذج الانحدار من متغير تابع قطعي يرمز له Y_i ، أما إذا تضمن النموذج عدة متغيرات مستقلة قطعية فيرمز لها X_{ij} حيث $i = 1,2,\cdots,n$ تشير إلى عدد المشاهدات بينما $j = 1,2,\cdots,p$ تشير إلى عدد المتغيرات وعندما تكون العلاقة بينهما خطية سنحصل على نموذج الانحدار الخطي له الشكل:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_P X_{iP} + \varepsilon_i \quad : \quad i = 1,2,\cdots,n \quad (13)$$

$\beta_0, \beta_1, \cdots, \beta_p$ معالم النموذج وهي قيم قطعية مجهولة يراد تقديرها وتسمى أيضاً بمعاملات الانحدار.

و ε_i:الأخطاء العشوائية وهي متغيرات عشوائية قطعية تتوزع وفق التوزيع الطبيعي وتكون مستقلة بمتوسط صفر وتباين ثابت σ^2 أي أن: $\varepsilon_i \sim N(0,\sigma^2) \;\; ; \;\; i = 1,2,\cdots,n$

وهنا نشير إلى أن: $\beta_0, \beta_1, \cdots, \beta_p$ و $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ هي مقادير مجهولة يجب العمل على تقديرها من بيانات العينة المسحوبة من المجتمع.

هناك منهجين لتقدير نماذج الانحدار هما: المنهج الكلاسيكي والمنهج البايزي حيث تعتبر معلمة التوزيع الاحتمالي للمجتمع مجهولة ثابتة ويتم تقديرها من خلال عينة عشوائية مسحوبة من هذا المجتمع و سنعرض في ما يلي المنهج البايزي . (Vittingghooff,Glidden,shiboski,2004)(Hastie,Tibshirani,friendman,2009) (إسماعيل،2001).

2-5-المنهج البايزي Bayesian Method:

في الاستدلال الإحصائي التقليدي يكون لدينا مجتمعاً إحصائياً موصوفاً بتوزيع احتمالي $\theta\epsilon\Theta$; $f(x,\theta)$ وتكون المعلمة المجهولة θ ثابتة ويراد تقديرها من خلال بيانات عينة عشوائية أما في الاستدلال البايزي يتم تفسير θ على أنها متغير عشوائي له الكثافة الاحتمالية $\pi(\theta)$ تعبر عما يعتقد الفرد حول قيمة حدوث θ قبل أي مشاهدة تؤخذ من العينة فسيكون سابقاً فإن منهج بايز يعتمد على دمج المعلومات الأولية مع معطيات العينة، حيث يتم تحويل الاحتمالات الأولية إلى احتمالات لاحقة، وبشكل عام يمثل الاحتمال الأولي درجة اعتقادنا المنطقية وهي مرتبطة دوماً بالوضعية الحالية لمعلوماتنا أو يصف خبرتنا السابقة حول المعلمة قبل الحصول على المعلمة ، أما الاحتمالات اللاحقة فهي تصف درجة اعتقادنا في القيم الممكنة للوسيط بعد الحصول على العينة العشوائية.

بالتالي سيكون لدينا رمزاً جديداً للتوزيع الاحتمالي الذي يصف المجتمع الإحصائي يصبح $f(x|\vartheta)$ ويعبر عن الاحتمال الشرطي لX علماً أن $\theta = \vartheta$ وسيكون$g(x,\theta)$ توزيع المشترك للمتغيرين θ, X ودالة المعقولية سيصبح لها الشكل:

$$L(x_1,x_2,\cdots,x_n|\theta)=\prod_{i=1}^{n}f(x_i|\theta) \qquad (14)$$

والآن ليكن لدينا $D_n=(X_1,X_2,\cdots,X_n)$عينة عشوائية مسحوبة من مجتمع إحصائي توزيعه الاحتمالي$f(D_n|\theta)$ ونرغب في تقدير θ باستخدام منهج بايز، سنقدم أولا بعض المفاهيم حول التقدير البايزي، (الصياد،1993)(Casella, Hogg,Robert, Allen, 2005)(Berger,2002).

2-5-1 تعريف التوزيع الاحتمالي القبلي prior Distribution:

إن التوزيع الاحتمالي القبلي $\pi(\theta)$ للمعلمة المجهولة θ هو توزيع احتمالي يصف كل المعلومات والخبرات عن المعلمة المجهولة قبل الحصول على العينة كما يصف درجة اعتقادنا عن القيم الممكنة لهذه المعلمة، (الصياد،1993).

2-5-2- تعريف التوزيع الاحتمالي البعدي Posterior Distribution:

لتكن لدينا $D_n=(X_1,X_2,\cdots,X_n)$ عينة عشوائية مسحوبة من مجتمع إحصائي توزيعه الاحتمالي $f(x|\vartheta)$ عندئذٍ التوزيع الاحتمالي البعدي $\pi(\theta|D_n)$ هو التوزيع الاحتمالي المشروط للمعلمة θ بشرط الحصول على العينة D_n وهو يصيف درجة اعتقادنا في القيم المختلفة للمعلمة θ بعد الحصول على العينة أي أن قراءات العينة تغير درجة اعتقادنا حول القيم المختلفة للمعلمة θ وذلك بتغيير التوزيع القبلي إلى توزيع البعدي، (الصياد،1993).

2-5-3-إيجاد التوزيع البعدي:

بفرض لدينا $D_n=(X_1,X_2,\cdots,X_n)$ عينة عشوائية مسحوبة من مجتمع إحصائي توزيعه الاحتمالي$f(x|\theta)$ نعلم أن دالة المعقولية

$$L(D_n|\theta)=f(D_n|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$$

التوزيع المشترك للمتغيرين D_n , θ يمكن كتابته بالشكل:

$$g(D_n,\theta)=\pi(\theta)f(D_n|\theta)=m(D_n)\pi(\theta|D_n)$$

حيث $m(D_n)$ التوزيع الهامشي للمتغير D_n وبالتالي فإن التوزيع البعدي للمعلمة θ بفرض أنها مستمرة له الشكل:

$$\pi(\theta|D_n)=\frac{g(D_n,\theta)}{m(D_n)}=\frac{\pi(\theta)f(D_n|\theta)}{m(D_n)} \tag{15}$$

وتوزيع الهامشي للمتغير D_n

$$m(D_n)=\int_{\theta} g(D_n,\theta)d\theta=\int_{\theta} \pi(\theta)f(D_n|\theta)d\theta \tag{16}$$

على أية حال ليس هناك حاجة لإيجاد $m(D_n)$ ، حيث التوزيع البعدي يمكن كتابته على الشكل، (الصياد،1993) :(Robert,2007)

$$\pi(\theta|D_n)\propto \pi(\theta)L(D_n|\theta) \tag{17}$$

2-5-3- تقدير بايز في حالة عدة معالم مجهولة:

بفرض لدينا مجموعة المعالم مستمرة $\underline{\theta}=(\vartheta_1,\vartheta_2,\cdots,\vartheta_J)$ ونريد تقديرها ولها التوزيع البعدي المشترك $\pi(\underline{\theta}|D_n)$ حيث D_n عينة المشاهدات فإن التوزيع الهامشي البعدي لأي معلمة مثلاً ϑ_j يمكن الحصول عليه من ايجاد تكامل $\pi(\underline{\theta}|D_n)$ على المعالم بعد تثبيت ϑ_j ويعطى التوزيع الهامشي البعدي للمعلمة ϑ_j

$$\pi_j(\vartheta_j|D_n)=\int \pi(\vartheta_1,\cdots,\vartheta_J\,|X)d\vartheta_1\cdots d\vartheta_{j-1}d\vartheta_{j+1}\cdots d\vartheta_J \tag{18}$$

بعد إيجاد التوزيع الهامشي لكل معلمة، يكون مقدر بايز للمعلمة ϑ_j وفق العلاقة:

$$\widehat{\vartheta_J}=\min_{\widehat{\vartheta_J}} E\left[l(\hat{\theta},\vartheta_j)\right]=\min_{\widehat{\vartheta_J}}\int l(\hat{\theta},\vartheta_j)\pi_j(\vartheta_j|x)d\vartheta_j \tag{19}$$

من الواضح أن التقديرات البايز تتطلب إجراء تكاملات تكون في بعض الأحيان معقدة يصعب من خلالها إيجاد التوزيعات البعدية أو يكون من الصعب إيجاد الصفات العددية للتوزيعات البعدية بشكل تحليلي بالأخص عندما نكون أمام توزيعات معقدة أو تقدير لعدة معالم مجهولة في وقت واحد ، و لذلك نقوم باستخدام سلسلة ماركوف مونت كارلو MCMC من خلال تطبيق بعض خوارزميات مثل: معاينة جيبس لتوليد عينات من التوزيعات البعدية المعقدة وإيجاد الصفات العددية لهذه التوزيعات، (Rowe,2003).

2-5-4-خوارزمية معاينة جيبس Gibbs sampler algorithm:

يعد منهج سلسلة ماركوف- مونت كارلو MCMC أحد أنواع المحاكاة التي تستخدم خواص سلسلة ماركوف للقيام بحسابات وتقديرات إحصائية معقدة. (Gilks, Richardson,Spiegelhalter,1995).

تمكن الباحثون مؤخراً من الاستفادة من سلسلة ماركوف-مونت كارلو من خلال عدة خوارزميات منها خوارزمية ميتروبوليس-هاستغ، معاينة جيبس، ميتروبوليس، معاينة المستقلة، لإيجاد التوزيعات البعدية والصفات العددية عندما يكون من الصعب إيجادها تحليلياً وقد أصبح منهج MCMC أكثر انتشاراً في حقول الإحصاء والإحصاء الفيزيائي وعلوم الحاسوب والعلوم المالية، حيث كانت خصائص MCMC موضع إهتمام العديد من الباحثين في الاستدلال البايزي ومعاينة التوزيع البعدي خصوصاً في الحالات التي يكون فيها النموذج معقداً.

ظهرت معاينة جيبس في البحث الذي نشره Geman&Geman عام 1984 في عمليات معالجة الصور وهي واحدة من أشهر الخوارزميات المستخدمة في الاستدلال البايزي في حالة المعالم المتعددة سنقوم بعرضها بالشكل التالي:

بفرض لدينا $D_n=\mathrm{X_1,X_2,\ldots,X_n}$ عينة عشوائية مسحوبة من مجتمع إحصائي توزيعه الاحتمالي $p(x|\underline{\theta})$ ولتكن $\underline{\theta}=(\vartheta_1,\vartheta_2,\cdots,\vartheta_k)$ مجموعة معالم مجهولة لها التوزيعات شرطية $\pi_j(\vartheta_j|\vartheta_1,\cdots,\vartheta_{j-1},\vartheta_{j+1},\cdots,\vartheta_k,\underline{x})$ حيث $j=$

$1,2,\cdots,k$ وهذه التوزيعات تدعى توزيعات شرطية تامة للمعالم $\vartheta_1,\vartheta_2,\cdots,\vartheta_k$ ، عندئذٍ يمكن وصف خطوات الخوارزمية، $\theta^{(t)}$ القيمة المولدة في تكرار t:

1− نفرض القيمة الابتدائية $\underline{\theta}^{(0)} = (\vartheta_1^{(0)},\vartheta_2^{(0)},\cdots,\vartheta_k^{(0)})$

2− نكرر الخطوات التالية من أجل $t = 1,2,\ldots,T$.

3− نولد قيمة المعلمة ϑ_1 ولتكن $\vartheta_1^{(t)}$ من التوزيع الشرطي $\pi_1\left(\vartheta_1 \middle| \vartheta_2^{(t-1)},\vartheta_3^{(t-1)},\cdots,\vartheta_k^{(t-1)},\underline{x}\right)$

4− نولد قيمة المعلمة ϑ_2 ولتكن $\vartheta_2^{(t)}$ من التوزيع الشرطي $\pi_2\left(\vartheta_2 \middle| \vartheta_1^{(t)},\vartheta_3^{(t-1)},\cdots,\vartheta_k^{(t-1)},\underline{x}\right)$

وهكذا ....نستمر حتى نولد قيمة المعلمة ϑ_k ولتكن $\vartheta_k^{(t)}$ من التوزيع الشرطي $\pi_k\left(\vartheta_k \middle| \vartheta_1^{(t)},\vartheta_2^{(t)},\cdots,\vartheta_{k-1}^{(t)},\underline{x}\right)$

بعد ذلك سنحصل على عينات مولدة من التوزيعات البعدية الهامشية لكل معلمة بعد إجراء عملية حرق نقوم بالتقديرات المطلوبة باستخدام الصفات العددية لتوزيع البعدي الهامشي لكل معلمة، (Ntzoufras,2011)(Robert,Casella,2013).

2−6− الانحدار الخطي الضبابي:

يهدف تحليل الانحدار الخطي التقليدي إلى تقدير دالة تمثل أفضل تلائم للبيانات المعطاة ويبنى هذا التحليل على العشوائية المعرفة وفق توزيع احتمالي تحت شرط أن البيانات قطعية وعلى رغم من وجود الكثير من التطبيقات التي استخدمت التحليل الانحدار الخطي إلا أنه تواجهنا عدة مشكلات:

1) أن يكون حجم العينة غير كافي (صغير) لإنجاز تحليل الانحدار التقليدي.

2) افتراضات التوزيع الطبيعي للأخطاء غير محققة.

3) العلاقة بين المخرجات والمدخلات غير محددة.

4) عدم معنوية النموذج.

5) افتراض خطية العلاقة غير ملائم.

6) عدم الدقة في البيانات والذي يعزى إلى أن الظاهرة المدروسة ذات طبيعة ضبابية حيث يهمل الانحدار التقليدي عدم الدقة بالبيانات.

في هذه الحالات الانحدار التقليدي لن يقدم أفضل ملائمة للبيانات المعطاة من أجل ذلك ظهرت الحاجة لتعميم الانحدار التقليدي الى الانحدار الخطي الضبابي وبحسب نوعية البيانات الإدخال والإخراج سنحصل على نموذج الانحدار الضبابي وفي هذا البحث سنركز على تحليل الانحدار الخطي الضبابي فيه مدخلات قطعية ومخرجات ضبابية (CIFO) (Crisp−Inputs and Fuzzy−Outputs).(Shapiro, 2005) .

2−6−1− نموذج الانحدار الخطي الضبابي مع البيانات (CIFO):

بفرض لدينا $(\boldsymbol{X_i}, y_i); i = 1,2,\ldots,n$ بيانات المشاهدات حيث $\boldsymbol{X_i} = [\mathbf{1}, x_{i1},\ldots,x_{im}]^T$ مدخلات ذات البعد (m + $1) \times 1$ و $y_i = (c_i, l_i, r_i)_T$ مخرجات أعداد ضبابية مثلثية، والغرض من الانحدار الخطي مع البيانات CIFO الحصول على دالة خطية معاملاتها ضبابية التي تمثل العلاقة بين مدخلات قطعية ومخرجات ضبابية ونموذج يعطى بالعلاقة التالية:

$$Y_i = A_0 + A_1X_{i1} + A_2X_{i2} + \cdots + A_mX_{im} = \boldsymbol{AX_i} \qquad (20)$$

حيث $\boldsymbol{A} = [A_0, A_1,\ldots,A_m]$ متجه المعاملات الضبابية المجهولة.

Y_i مقدر الخرج الضبابي باستخدام المدخلات قطعية .

تتحول مشكلة الانحدار الخطي الضبابي لإيجاد المعاملات A المثلى لكي تكون القيم المقدرة Y_i قريبة من قيم المشاهدات y_i.(Wang,etal,2020)(Alstolanty,Alnagash,2015)

2-6-2- طرائق تقدير نموذج الانحدار الخطي الضبابي مع البيانات (CIFO):

بفرض لدينا نموذج الانحدار الخطي الضبابي مع البيانات (CIFO) ولتكن $A_j = (\alpha_j,\ \beta_j, \gamma_j)_T\ ; j = 0,1,2, \dots, m$ عدد مثلثي ضبابي حيث α_j المركز و β_j الانتشار الايسر و γ_i الانتشار الأيمن ولدينا $X_{i0} = 1$ من أجل أي $i = 1,2, \dots, n$ وسنفرض أن المدخلات جميعها موجبة أي أن $X_{ij} > 0$ من أجل كل قيم i, j بالاعتماد على تعريف العمليات على الأعداد الضبابية LR نحصل على:

$$Y_i = \sum_{j=0}^{m} A_j X_{ij} = (\sum_{j=0}^{m} \alpha_j X_{ij}\, , \sum_{j=0}^{m} \beta_j X_{ij}\, , \sum_{j=0}^{m} \gamma_j X_{ij})_T \qquad (21)$$

.(Wang,etal,2020)(Alstolanty,Alnagash,2015)

سنعرض فيما يلي ثلاثة طرائق لتقدير معاملات الانحدار الخطي الضبابي مع البيانات (CIFO)

2-6-2-1- أولاً: نموذج الإمكانية لتاناكا:Tanaka possibilistic model

بفرض أن المعاملات في المعادلة (21) أعداد ضبابية مثلثية ضبابية متناظرة أي $\beta_j = \gamma_j = e_j$ وبالتالي: $A_j = (\alpha_j, e_j)_T\ ; j = 0,1,2, \dots, m$ و الحد الأدنى لملائمة نموذج H لكل البيانات المعطاة تعطى من قبل صانع القرار وهدف النموذج أن يقلل المجموع الكلي للانتشار المعاملات الضبابية بشرط $\bar{h}_i \geq H$ من ثم توظف مسألة (LP) لتقدير معلمات الانحدار الضبابي، والتي تبنى على اساس دالة الهدف والقيود كما يلي:

$$\begin{aligned}
&\min_{\alpha_j, e_j} J = e_0 + e_1 + \cdots + e_m \\
s.t \quad &\sum_{j=0}^{m} \alpha_j X_{ij} + (1 - H) \sum_{j=0}^{m} e_j X_{ij} \geq c_i + (1 - H) r_i \\
&-\sum_{j=0}^{m} \alpha_j X_{ij} + (1 - H) \sum_{j=0}^{m} e_j X_{ij} \geq -c_i + (1 - H) l_i \qquad (22) \\
&e_j \geq 0\ ; i = 1,2, \dots, n \qquad j = 0,1,2, \dots, m
\end{aligned}$$

2-6-2-2- ثانياً: نموذج المربعات الصغرى الضبابية:

تستخدم مبدأ المربعات الصغرى في إيجاد المعاملات في النموذج (21) من خلال تقليل المسافة الكلية بين المخرجات المقدرة الضبابية والمخرجات المشاهدات الضبابية واستخدمت عدة معاير لقياس المسافة من أهمها:

1- نموذج Diamond المربعات الصغرى الضبابية

قدم من قبل (Diamond ,1988) استخدم المساقة التربيعية ويعطى بالشكل التالي:

$$\min_{\alpha_j, e_j, \gamma_j} D^2 = \sum_{i=1}^{n} (y_i - Y_i)^2$$
$$= \sum_{i=1}^{n} \left(\left(c_i - \sum_{j=0}^{m} \alpha_j X_{ij} \right)^2 + \left(l_i - \sum_{j=0}^{m} \beta_j X_{ij} \right)^2 + \left(r_i - \sum_{j=0}^{m} \gamma_j X_{ij} \right)^2 \right) \quad (23)$$
$$s.t. \quad \beta_j \geq 0 \,, \gamma_j \geq 0 \,; i = 1,2, \dots, n \qquad j = 0,1,2, \dots, m$$

2- نموذج Zeng القيمة المطلقة الأقل:

استخدم نموذج Zeng المسافة المطلقة بين المخرجات المقدرة الضبابية والمخرجات المشاهدات الضبابية ويعطى بالشكل:

$$\min_{\alpha_j, e_j, \gamma_j} d = \sum_{i=1}^{n} |y_i - Y_i|$$
$$= \sum_{i=1}^{n} \left(\left| c_i - \sum_{j=0}^{m} \alpha_j X_{ij} \right| + \left| l_i - \sum_{j=0}^{m} \beta_j X_{ij} \right| + \left| r_i - \sum_{j=0}^{m} \gamma_j X_{ij} \right| \right) \quad (24)$$
$$s.t. \quad \beta_j \geq 0 \,, \gamma_j \geq 0 \,; i = 1,2, \dots, n \qquad j = 0,1,2, \dots, m$$

.(Shapiro,2005)(Alstolanty,Alnagash,2015)

2-6-2-3- ثالثاً: النموذج البايزي لتقدير نموذج الانحدار الخطي الضبابي:

بفرض لدينا $(\boldsymbol{X_i}, y_i); i = 1,2, \dots, n$ بيانات المشاهدات حيث $\boldsymbol{X_i} = [\boldsymbol{1}, x_{i1}, \dots, Tx_{im}]^{\boldsymbol{T}}$ مدخلات ذات البعد m+1 و $y_i = (c_i, r_i)_T$ مخرجات أعداد ضبابية مثلثية متناظرة، ونموذج الخطي الضبابي مع الخطأ العشوائي الضبابي يعطى بالعلاقة التالية:

$$Y_i = A_0 + A_1 X_{i1} + A_2 X_{i2} + \cdots + A_m X_{im} + E_i \quad (25)$$

حيث معاملات أعداد ضبابية مثلثية ضبابية متناظرة $A_j = (\alpha_j, e_j)_T \,; j = 0,1,2, \dots, m$ وليكن:

$\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_m]$ و $\boldsymbol{e} = [e_0, e_1, \dots, e_m]$ رلكز وانتشارات المعاملات الضبابية،(Stein, (Wang,etal,202)(Beer,Kreinovich,2013 .

من خلال دراستنا لاحظنا عدة معالجات للموضوع منها التقدير على مرحلتين، في المرحلة الأولى يتم تقدير المراكز $\boldsymbol{\alpha}$ باستخدام المربعات الصغرى العادية بعد إزالة الضبابية عن المخرجات، في المرحلة الثانية لإيجاد الانتشارات $\boldsymbol{e}$ يستخدم منهج بايز بتطبيق خوارزمية الرفض ABC. بافتراض أن $\boldsymbol{e}$ متغير عشوائي له توزيع قبلي $\pi(\boldsymbol{e})$ وسيحدد التوزيع البعدي $\pi(\boldsymbol{e}|\boldsymbol{y})$ حيث $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^t$ المشاهدات من العلاقة التالية:

$$\pi(\boldsymbol{e}|\boldsymbol{y}) \propto l(\boldsymbol{e},\boldsymbol{y})\pi(\boldsymbol{e})$$

حيث $l(\boldsymbol{e},\boldsymbol{y})$ دالة المعقولية العظمى، (Stein, Beer,Kreinovich,2013)(Wang,etal,202) .

في هذا البحث نستخدم منهج بايز لتقدير هذه المعالم وسنفرض أن $\boldsymbol{\alpha}$ و $\boldsymbol{e}$ متغيران عشوائيان لكل منهما توزيع احتمالي قبلي $\pi_2(\boldsymbol{e}),\pi_1(\boldsymbol{\alpha})$ سيكون لدينا التوزيع البعدي $\pi(\boldsymbol{\alpha},\boldsymbol{e}|\boldsymbol{y})$ حيث $\boldsymbol{y}=[y_1,y_2,\dots,y_n]^t$ المشاهدات نوجد التوزيع البعدي من العلاقة التالية:

$$\pi(\boldsymbol{\alpha},\boldsymbol{e}|\boldsymbol{y}) \propto l(\boldsymbol{\alpha},\boldsymbol{e},\boldsymbol{y})\pi_1(\boldsymbol{\alpha})\pi_2(\boldsymbol{e}) \qquad (26)$$

حيث $l(\boldsymbol{\alpha},\boldsymbol{e},\boldsymbol{y})$ دالة الإمكانية العظمى.

عند تطبيق المنهج البايزي نواجه مشكلتان أساسيات **أولاً:** لا يوجد بشكل عام تعريف مقبول لدالة الإمكانية العظمى يمكن تطبيقه بشكل عملي، **ثانياً:** تحديد التوزيعات القبلية في حالة المعالم الضبابية وسنعرض فيما يلي خوارزميتان لتوليد عينات من التوزيع البعدي خالية من دالة المعقولية ونقترح خوارزمية تدمج بين معاينة جيبس وخوارزمية ABC ونقدم حلاً لإيجاد تقدير معالم نموذج الانحدار الخطي الضبابي باستخدام الخوارزمية المقترحة.

2-6-2-3-1- خوارزمية معاينة الرفض :Rejection sampling algorithm

بشكل العام لا يوجد تعريف مقبول لدالة الإمكانية العظمى لنموذج ذي معاملات ضبابية يمكن تطبيقها بشكل عملي لذلك خوارزميات MCMC بشكلها التقليدي ليست قابلة للتطبيق لإيجاد عينات من التوزيع البعدي، توفر لنا خوارزمية الرفض حل خالٍ من دالة الإمكان.

بفرض لدينا معلمات θ لها التوزيع القبلي $\pi(\theta)$ ولها توزيع قبلي و $(\boldsymbol{X_i},y_i);i=1,2,\dots,n$ بيانات المشاهدات بحيث $\boldsymbol{X_i}$ المدخلات و y_i المخرجات ولها نموذج $M(\theta,\boldsymbol{X_i})$ فإذا استطعنا توليد قيمة من قيم θ يكون لدينا

$Y_i=M(\theta,\boldsymbol{X_i})\,;i=1,2,\dots,n$ حيث $\boldsymbol{Y_\theta}=[Y_1,Y_2,\dots,Y_n]^t$

تقدم خوارزمية الرفض تحويل العينات من التوزيع القبلي إلى عينات من التوزيع البعدي من خلال الاحتفاظ بكل قيمة مولدة إذا كانت البيانات المحسوبة $\boldsymbol{Y_\theta}$ تساوي البيانات المشاهدة للخوارزمية الخطوات التالية:

1- نكرر الخطوات التالية من أجل t=1,2,…,T .

2- نكرر الخطوات 3 ، 4.

3- { نولد قيمة θ من توزيع احتمالي قبلي $\pi(\theta)$.

4- نحسب قيمة $\boldsymbol{Y_\theta}$ من نموذج $Y_i=M(\theta_1,\boldsymbol{X_i})$. }

5- حتى نحصل على $\boldsymbol{Y_\theta}=y$.

6- $\theta\to\theta^{(t)}$. سنحصل على عينة بحجم T

سيتولد لدينا عينات من التوزيع البعدي ولكن احتمال تساوي القيم المولدة مع البيانات المشاهدة شبه مستحيل ولا يمكن تطبيقه بشكل عملي بسبب نسبة القبول المنخفضة جداً والتعقيد الحسابي لذلك ظهرت الحاجة لتعديل شرط القبول لنكون أمام خوارزمية الرفض ABC (حسابات بايز التقريبية) ،(Wang,etal,2020) (Lintusaari,Gutmann,Dutta, Kaski, Corander, 2017).

2-6-2-3-2- خوارزمية الرفض ABC (حسابات بايز التقريبية):

rejection ABC (approximate Bayesian computation) algorithm

تولد خوارزمية الرفض عينات من التوزيع البعدي ولكن شرط تساوي القيم المولدة مع البيانات المشاهدة شبه مستحيل ولجعل التوليد معقول عملياً نستبدل معيار القبول في الخوارزمية $\boldsymbol{Y}_\theta = y$ لتخفيفه إلى الشرط إلى $D(\boldsymbol{Y}_\theta, y) \leq \epsilon$ حيث $\epsilon \geq 0$ و D دالة المسافة لقياس الاختلاف بين المشاهدات و البيانات المولدة وبهذا المعيار لا نحصل على بيانات مولدة من التوزيع البعدي وإنما قيم مولدة من توزيع تقريبي للتوزيع البعدي:

$$\pi_{D,\epsilon}(\theta|\boldsymbol{y}) \propto P(D(\boldsymbol{Y}_\theta, y) \leq \epsilon)\pi(\boldsymbol{e})$$

الذي هو توزيع البعدي للمعلمة بشرط وقوع الحدث $D(\boldsymbol{Y}_\theta, y) \leq \epsilon$ سنحصل على قيم تقريبية لعينة من التوزيع البعدي ، ويكون الخوارزمية لها الخطوات التالية:

1- نكرر الخطوات التالية من أجل t=1,2,…,T.

2- نكرر الخطوات 3 ، 4.

3- { نولد قيمة θ من توزيع احتمالي قبلي $\pi(\theta)$.

4- نحسب قيمة $\boldsymbol{Y}_\theta$ من نموذج $Y_i = M(\theta_1, \boldsymbol{X_i})$. }

5- حتى نحصل على $D(\boldsymbol{Y}_\theta, y) \leq \epsilon$.

6- $\theta \to \theta^{(t)}$. سنحصل على عينة بحجم T،(Wang,etal,2020) (Lintusaari,Gutmann,Dutta, Kaski, Corander, 2017).

2-6-2-3-3- الخوارزمية المقترحة:

بفرض لدينا $(\boldsymbol{X_i}, y_i); i = 1,2, \dots, n$ بيانات المشاهدات و $y_i = (c_i, r_i)_T$ مخرجات أعداد ضبابية مثلثية متناظرة، ونموذج الخطي الضبابي مع الخطأ العشوائي الضبابي يعطى بالعلاقة التالية:

$$Y_i = M(\boldsymbol{\alpha}, \boldsymbol{e}, \boldsymbol{X_i}) = A_0 + A_1X_{i1} + A_2X_{i2} + \cdots + A_mX_{im} + E_i$$

حيث المعاملات أعداد ضبابية مثلثية ضبابية متناظرة $A_j = (\alpha_j, e_j)_T\ ; j = 0,1,2, \dots, m$ وليكن $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_m]$ و $\boldsymbol{e} = [e_0, e_1, \dots, e_m]$ مراكز وانتشارات المعاملات الضبابية نستخدم منهج بايز لتقدير المعالم سنفرض أن $\boldsymbol{\alpha}$ و $\boldsymbol{e}$ متغيران عشوائيان لكل منهما توزيع احتمالي قبلي $\pi_2(\boldsymbol{e}), \pi_1(\boldsymbol{\alpha})$ سيكون لدينا التوزيع البعدي $\pi(\boldsymbol{\alpha}, \boldsymbol{e}|\boldsymbol{y})$ حيث $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^t$ المشاهدات نوجد التوزيع البعدي من العلاقة التالية:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{e}|\boldsymbol{y}) \propto l(\boldsymbol{\alpha}, \boldsymbol{e}, \boldsymbol{y})\pi_1(\boldsymbol{\alpha})\pi_2(\boldsymbol{e})$$

من خلال دمج خوارزمية جيبس مع خوارزمية حسابات بايز التقريبية نقدم حلاً يعتمد على منهجية بايز لتقدير معاملات النموذج الضبابي عند تطبيق خوارزمية المقترحة ظهرت لدينا مشكلتين أساسيتين أولا شكل حسابي لدالة الإمكانية ولحلها استخدامنا خوارزمية حسابات بايز التقريبية وثانيا تحديد التوزيعات القبلية من أجل ذلك افترضنا توزيعات منتظمة محددة بالإمكانيات الممكنة لكل من الانتشارات و المراكز $e_j \sim U(0, ke_j^*)\ ; j = 0,1,2, \dots, m$ حيث e_j^* قيم ابتدائية تحدد من خلال طرائق الإمكانية و k عدد موجب وتوزيع القبلي للمراكز $\alpha_j \sim U(\alpha_j^* - e_j^*, \alpha_j^* + e_j^*)$ حيث α_j^* قيم ابتدائية تحدد من خلال طريقة المربعات الصغرى وتكون خطوات الخوارزمية المقترحة بالشكل التالي:

1- نقدر $\boldsymbol{\alpha}^{(0)}$ المراكز من خلال طريقة Zeng

2- نقدر $\boldsymbol{e}^{(0)}$ الانتشارات من خلال طريقة Tanaka

3- نكرر الخطوات التالية من أجل $t = 1,2, \dots, T$

4- نولد قيمة المعلمة $\boldsymbol{e} = [e_0, e_1, \dots, e_m]$ ولتكن $\boldsymbol{e}^{(t)} = \left[e_0^{(t)}, e_1^{(t)}, \dots, e_m^{(t)}\right]$ من $\pi_1\left(\boldsymbol{e} \middle| \alpha^{(t-1)}, \boldsymbol{y}, \underline{x}\right)$ باستخدام خوارزمية الرفض ABC كما يلي:

a) نكرر الخطوات b، c

b) { نولد القيم $e_j^{(t)}; j = 0,1, \dots, m$ من توزيعات احتمالية قبلية $e_j \sim U\left(0, ke_j^{(t-1)}\right)$; $j = 0,1, \dots, m$

c) نحسب قيمة $\boldsymbol{Y}_{\alpha^{(t-1)}, \boldsymbol{e}^{(t)}}$ من نموذج $Y_{i,\alpha^{(t-1)},\boldsymbol{e}^{(t)}} = M(\alpha^{(t-1)}, \boldsymbol{e}^{(t)}, \boldsymbol{X_i})$ }

d) حتى نحصل على $\sum_{i=1}^{n} \frac{1}{n} D\left(Y_{i,\boldsymbol{e}^{(t)},\alpha^{(t-1)}}, y_i\right) \le \epsilon$

5- نولد قيمة المعلمة $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_m]$ ولتكن $\boldsymbol{\alpha^{(t)}} = \left[\alpha_0^{(t)}, \alpha_1^{(t)}, \dots, \alpha_m^{(t)}\right]$ من $\pi_2\left(\boldsymbol{\alpha} \middle| \boldsymbol{e}^{(t)}, \boldsymbol{y}, \underline{x}\right)$ باستخدام خوارزمية الرفض ABC

a) نكرر الخطوات b، c

b) { نولد القيم $\alpha_j^{(t)}; j = 0,1, \dots, m$ من توزيع احتمالي قبلي $\alpha_j \sim U\left(\alpha_j^{(t-1)} - e_j^{(t)}, \alpha_j^{(t-1)} + e_j^{(t)}\right)$; $j = 0,1, \dots, m$

c) نحسب قيمة $\boldsymbol{Y}_{\boldsymbol{e}^{(t)}, \boldsymbol{\alpha}^{(t)}}$ من نموذج $Y_{i,\boldsymbol{e}^{(t)},\boldsymbol{\alpha}^{(t)}} = M(e_j, \boldsymbol{X_i})$ }

d) حتى نحصل على $\sum_{i=1}^{n} \frac{1}{n} D\left(Y_{i,\boldsymbol{e}^{(t)},\boldsymbol{\alpha}^{(t)}}, y_i\right) \le \epsilon$

6) $(\boldsymbol{e}, \boldsymbol{\alpha}) \to (\boldsymbol{e}^{(t)}, \boldsymbol{\alpha}^{(t)})$ عينة بحجم T المولدة من التوزيع الشرطي التقريبي $\pi_{D,\epsilon}(\boldsymbol{e}, \boldsymbol{\alpha} | \boldsymbol{y})$

بعد ذلك نوجد التقديرات من خلال العينة المولدة واستخدامنا متوسط العينة المولدة للحصول على تقديرات المطلوبة واستخدمنا متوسط مسافة Diamond كمعيار لقياس الاختلاف بين المشاهدات والبيانات المولدة.

من خلال الخوارزمية المقترحة نقدم حلاً عددياً لتقديرات بايز التي تأخذ عشوائية الظواهر المدروسة بالحسبان ولها قدرة عالية على التنبؤ ويحقق الموازنة في التقديرات بين منهجي الإمكانية و المربعات الصغرى.

وتم اختبار الخوارزمية المقترحة من خلال بيانات مولدة ومن ثم التطبيق على بيانات واقعية.

3- الجانب التطبيقي:

في هذا القسم قمنا بالاختبار الخوارزمية المقترحة على بيانات مولدة وبيانات واقعية ومقارنتها مع طرائق الإمكانية نموذج تاناكا و الحد الأدنى للملائمة المفترض $H = 0$ وطرائق المربعات الصغرى نموذج Zeng.

مثال1: تم بناء النموذج الانحدار الخطي الضبابي باستخدام ثلاث متغيرات إدخال لها توزيعات $X_1, X_2 \sim N(2,1)$ و $X_3 \sim N(4,1)$ والخطأ العشوائي $E_i \sim N(0,1)$ وأوجدنا المراكز المخرجات (c_i) بتوليد قيم المتغيرات X_1, X_2, X_3, E_i ومن خلال العلاقة التالية:

$$c_\text{i} = 2X_1 + 3X_2 + 2X_3 + E_i$$

و الانتشارات تم توليدها من التوزيع $r_\text{i} \sim N(4,1)$

وكانت البيانات كما في الجدول التالي:

الجدول رقم (1): البيانات المولدة

X_3	X_2	X_1	$Y_i = (c_i, r_i)_T$	n
4.264610	3.8845276	1.73805889	$(22.80725,\ 3.366877)_T$	1
1.896503	2.9135196	3.61841889	$(18.71452,\ 3.421943)_T$	2
4.850253	1.9874969	1.08033228	$(16.91136,\ 3.617333)_T$	3
2.868136	3.0128176	0.48396665	$(16.09859,\ 4.506642)_T$	4
2.983301	1.1436337	2.02799804	$(14.58889,\ 4.795787)_T$	5
3.307440	2.4561162	3.45999798	$(20.38313,\ 6.382205)_T$	6
4.302351	-0.3221226	2.04982068	$(11.37544,\ 3.647870)_T$	7
4.225561	2.3508017	4.48175769	$(24.03057,\ 4.607913)_T$	8
2.83081	0.7344797	1.42281158	$(12.46551,\ 4.550187)_T$	9
6.264924	-0.1397076	0.04866684	$(13.64247,\ 4.704391)_T$	10
3.092505	0.3676157	1.218161	$(10.31607,\ 5.028792)_T$	11
3.036732	2.218747	1.995732	$(16.82465,\ 4.015910)_T$	12
2.035748	2.1607081	1.674776	$(14.36478,\ 3.725369)_T$	13
4.9919	2.3732137	3.561849	$(24.66938,\ 1.773772)_T$	14
5.003484	-0.2233581	1.60289	$(12.14366,\ 4.709382)_T$	15
5.167166	2.6226804	2.981529	$(23.58879,\ 3.471180)_T$	16
4.297939	3.1072833	2.729842	$(24.30680,5.094771)_T$	17
4.926146	1.1761818	2.816123	$(19.12939,\ 3.424199)_T$	18
3.99903	3.2737475	2.709298	$(22.43574,\ 3.099474)_T$	19
3.233092	0.570221	2.341204	$(11.94261,\ 2.233733)_T$	20

المصدر: من إعداد الباحث اعتماداً على اللغة البرمجية R

تم إيجاد المقدرات باستخدام طرائق الإمكانية (نموذج Tanaka) و طرائق المربعات الصغرى (نموذج Zeng) ومن ثم تطبيق الخوارزمية المقترحة أولاً من أجل عينة بحجم 10 ثم بالاستخدام عينة بحجم عشرين من أجل الحكم على جودة أداء كل طريقة بغرض المقارنة بين طرائق السابقة تم استخدام مقياس لجودة التلاءم(GOF) بالاعتماد على مسافة Diamond التي تعرف كما يلي، (Wang,Zhang,Mei,2007):

$$GOF = \frac{1}{n}\sum_{i=1}^{n}(c_1 - l_1 - c_2 + l_2)^2 + (c_1 - c_2)^2 + (c_1 + r_1 - c_2 - r_2)^2 \qquad (27)$$

ذلك من خلال كتابة برنامج نصي باللغة البرمجية R (لغة برمجية إحصائية)، حيث تستطيع هذه اللغة القيام بالعديد من عمليات تحليل البيانات بحيث يتم تنظيم هذه التحليلات ضمن ما يسمى بالحزم Packages مما يعني قدرة الباحثين على تطوير البرامج المختلفة الأمر الذي ساهم بانتشار استخدامها في المجالات الأكاديمية (Cotton، 2013).

عند تطبيق االخوارزمية المقترحة استخدمنا الحد الأدنى $\epsilon = 3.1$ من أجل حجم عينة 10 و $\epsilon = 3.75$ من أجل حجم عينة 20 وتكرار $T = 20000$ في الحالتين وكانت النتائج بالجدول التالي:

الجدول رقم (2): نتائج تطبيق الطرائق المستخدمة في تقدير معالم نموذج الضبابي على البيانات المولدة

n	A_3	A_2	A_1	A_0	GOF	n
Tanaka	$(1.66,0.24)_T$	$(2.34,0.12)_T$	$(1.63,0.61)_T$	$(3.46,3.13)_T$	6.886098	10
Zeng	$(1.78,0.05)_T$	$(2.47,0)_T$	$(1.63,0.005)_T$	$(2.72,4.33)_T$	3.083984	
نموذج المقترح	$(1.73, 0.058)_T$	$(2.57, 0.02)_T$	$(1.66, 0.09)_T$	$(2.53,3.89)_T$	2.775453	
الإمكانية	$(1.85,0.13)_T$	$(2.99,0)_T$	$(1.38,0.31)_T$	$(2.08,4.95)_T$	11.99794	20
Zeng	$(1.9,0)_T$	$(2.81,0)_T$	$(1.68,0)_T$	$(1.33,4.01)_T$	3.739739	
نموذج المقترح	$(1.93,0.004)_T$	$(2.84,0.003)_T$	$(1.74,0.003)_T$	$(1.11,3.97)_T$	3.69926	

المصدر: من إعداد الباحث اعتماداً على مخرجات اللغة البرمجية R

نلاحظ من الجدول رقم (2) أن قيم GOF في التقديرات المعتمدة على الخوارزمية المقترحة أصغر من مقابلاتها في طرائق الإمكانية وطرائق المربعات الصغرى من أجل حجمي العينة مختلفين 10 و 20.

مثال2(بيانات واقعية):

استخدمنا البيانات التي تمثل GDP الناتج المحلي الإجمالي (Y_i) المخرجات، أما المدخلات فهي (X_1) وعدد العاملين (X_2) في اليابان، وشكلت المخرجات الضبابية $Y_i = (c_i, r_i)_T$ بتحديد نسبة 5% من المخرجات c_i كانتشار لكل خرج وكما تم افتراض أن المخرجات أعداد ضبابية مثلثية متناظرة و البيانات في الجدول التالي، (Lee, Tanaka, 1999):

الجدول رقم (3): البيانات الواقعية GDP الناتج المحلي الإجمالي

n	GDP $Y_i = (c_i, r_i)_T$	عدد العاملين X_2	الدخل X_1
1	$(124.5,6.225)_T$	137	102.5
2	$(129.4,6.470)_T$	138.1	103.5
3	$(135.1,6.755)_T$	141.5	104.8
4	$(142.3,7.115)_T$	144.8	106.1
5	$(150.1,7.505)_T$	148.1	107.5
6	$(154.3,7.715)_T$	146	108.7
7	$(159.2,7.960)_T$	148.8	109.5
8	$(164.0,8.200)_T$	151	110.7،
9	$(167.9,8.395)_T$	151.4	112.5
10	$(174.4,8.720)_T$	153.7	113.2
11	$(182.1,9.105)_T$	155.6	114
12	$(187.4,9.370)_T$	158.8	114.9
13	$(195.2,9.760)_T$	161.9	116
14	$(207.3,10.365)_T$	166.1	118
15	$(217.3,10.865)_T$	169.1	120.3
16	$(228.3,11.415)_T$	173	122.6
17	$(237.0,11.850)_T$	176	125
18	$(239.4,11.970)_T$	174.8	126.3

المصدر: (Lee, Tanaka, 1999)

وقدرنا المعالم الضبابية المثلثية باستخدام طرائق الإمكانية (نموذج تاناكا) و طرائق المربعات الصغرى (نموذج Zeng) ومن ثم تطبيق الخوارزمية المقترحة استخدمنا حد الأدنى $\epsilon = 8.8$ وتكرار الخوارزمية $T = 20000$ وكانت النتائج بالجدول التالي

الجدول رقم (4): نتائج تطبيق الطرائق المستخدمة في تقدير معالم نموذج الضبابي على البيانات واقعية

n	A_2	A_1	A_0	GOF
Tanaka	$(2.27,0)_T$	$(1.6,0.07)_T$	$(-328.88\ ,0)_T$	18.38411
Zeng	$(\ 2.17,0)_T$	$(1.68,0.56)_T$	$(\ -330.6,0)_T$	8.814696
نموذج المقترح	$(2.3, 0.0001)_T$	$(1.61,0.057)_T$	$(-333.94, 0.003)_T$	8.289659

المصدر: من إعداد الباحث اعتماداً على مخرجات اللغة البرمجية R

نلاحظ من الجدول رقم (4) أن قيمة GOF في التقديرات باعتماد على الخوارزمية المقترحة أصغر من مقابلاتها في طرائق الإمكانية وطرائق المربعات الصغرى.

4-الاستنتاجات والتوصيات:

4-1-الاستنتاجات: Conclusions

اعتماداً على ما سبق تم الحصول على مقدرات معالم نموذج الانحدار الخطي الضبابي باستخدام طرائق الإمكانية وطرائق المربعات الصغرى والخوارزمية المقترحة وتم اختبار دقة النتائج باستخدام أسلوب المحاكاة ومن ثم تم التطبيق على بيانات واقعية وتوصلنا من خلال ذلك إلى النتائج التالية:

1) قدمت الخوارزمية المقترحة حلاً عددياً لتقدير النموذج الانحدار الضبابي لا يعتمد على الأساليب الحتمية (حساب الأمثلية) بل يعتمد على منهج بايز بالتقدير الذي يأخذ عشوائية الظواهر المدروسة بالحسبان.
2) تحقق الخوارزمية المقترحة التوازن بين طرائق الإمكانية وطرائق المربعات الصغرى من خلال الافتراض القيم الابتدائية لكل من الانتشارات والمراكز.
3) تقدم الخوارزمية المقترحة حلاً خالٍ من دالة الإمكانية مع ملاحظة أن التعقيد الحسابي للخوارزمية يزداد كلما قلت قيمة ϵ وتقل نسبة القبول وبالعكس إذا كانت قيمة ϵ كبيرة ترتفع نسبة القبول لذلك عند التطبيق العلمي استخدمنا قيم قريبة من أخطاء طرائق المربعات الصغرى وحصلنا على نتائج أفضل.
4) يتضح من النتائج أن قيم GOF في التقديرات باعتماد على الخوارزمية المقترحة أصغر من تلك التي تقابلها في طرائق الإمكانية وطرائق المربعات الصغرى، سواء عند استخدام بيانات مولدة أو عند استخدام بيانات واقعية.

4-2-التوصيات: Recommendations

1- مما سبق نوصي باستخدام الخوارزمية المقترحة كونها أعطت نتائج أفضل من طريقة Zeng و طريقة Tanaka .
2- نوصي بتطبيق الخوارزمية المقترحة باستخدام معايير جديدة لقياس الاختلاف بين البيانات الضبابية أو المزج بين عدة معايير للحصول على معيار جديد.
3- تطبيق الخوارزمية المقترحة باستخدام توزيعات قبلية طبيعية.
4- تعميم الخوارزمية المقترحة لتقدم حلاً لنماذج انحدار ضبابية لا خطية.
5- تطبيق الخوارزمية من أجل معاملات أعداد ضبابية مثلثية بحالة العامة.

5-References:

1) **Alsoltany, S. N., & Alnaqash, I. A.** (2015). Estimating Fuzzy Linear Regression Model for Air Pollution Predictions in Baghdad City. Al-Nahrain Journal of Science, 18(2), 157-166.
2) **Barros, L. C. D., Bassanezi, R. C., & Lodwick, W. A.** (2017). A first course in fuzzy logic, fuzzy dynamical systems, and biomathematics: theory and applications.
3) **Casella, G., & Berger, R. L.** (2002). Statistical inference (Vol. 2, pp. 337-472). Pacific Grove, CA: Duxbury
4) **Cotton, R.,** (2013). Learning R, O'Reilly Media, Inc., United States of America, 377
5) **Fullér, R.** (1995). Neural fuzzy systems.
6) **HASTIE T., TIBSHIRANI R. and FRIEDMAN J.,** 2009- The Elements of Statistical Learning Data Mining. Inference, and Prediction. Springer, 2th ED, Berlin, 764
7) **Hogg, Robert. V., Allen. T .Craig,** (2005) - Introduction to Mathematical Statistics, Macmillan Publishing, Co., Inc., New York
8) **J. Lintusaari, M.U. Gutmann, R. Dutta, S. Kaski, J. Corander,** (2017) .Fundamental and recent developments in approximate Bayesian computation, Syst. Biol.66 (1) e66–e82
9) **L.A. Zadeh**, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
10) **Lee, H. and Tanaka, H.** (1999) Fuzzy approximations with non-symmetric fuzzy parameters in fuzzy regression analysis. Journal of the Operations Research Society Japan 42: 98-112
11) **M. Stein, M. Beer, V. Kreinovich,** (2013). Bayesian approach for inconsistent information, Inform. Sci. 245 96–111
12) **Nguyen, H. T., Walker, C. L., & Walker, E. A.** (2018). A first course in fuzzy logic. CRC press.
13) **Ntzoufras, I.** (2011). Bayesian modeling using WinBUGS (Vol. 698). John Wiley & Sons, Inc. Hoboken, New Jersey
14) **Robert, C.** (2007). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media
15) **Robert, C., & Casella, G.** (2013). Monte Carlo statistical methods. Springer Science & Business Media..
16) **Rowe, D. B.** (2003). Multivariate Bayesian statistics: models for source separation and signal unmixing. Chapman and Hall/CRC

17) **Shapiro, A. F.** (2005). Fuzzy regression models. Article of Penn State University, 102(2), 373-383

18) **Viertl, R.** (2011). Statistical methods for fuzzy data. John Wiley & Sons.

19) **Vittingghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E.** (2004). Regression Methods in Biostatistics. 2004. San Fransico: Springer.

20) **Wang N., Zhang W.-X., Mei C.-L.** (2007) Fuzzy nonparametric regression based on local linear smoothing technique. *Information Sciences* 177: 3882-3900

21) **Wang, N., Reformat, M., Yao, W., Zhao, Y., & Chen, X.** (2020). Fuzzy Linear regression based on approximate Bayesian computation. Applied Soft Computing, 97, 106763.

22) **Zimmermann, H. J.** (2011). Fuzzy set theory—and its applications. Springer Science & Business Media.

23) **Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). Markov chain Monte Carlo in practice. Chapman and Hall/CRC.UK**

24) **اسماعيل، محمد عبد الرحمن،** (2001). تحليل الانحدار الخطي. الإدارة العامة للنشر بمعهد الإدارة العامة، المملكة العربية السعودية، 495 صفحة.

25) **د. الصياد، جلال مصطفى،** 1993 – الاستدلال الإحصائي. الطبعة الثانية، دار المريخ للنشر، الرياض، المملكة العربية السعودية،657 صفحة.

26) **زينه، محمّد بشر**.(2019)." توظيف المجموعات الضّبابيّة في توسيع نظريّة صفوف الانتظار وبعض تطبيقاتها"، رسالة دكتوراه في الإحصاء الرياضي، كلية العلوم ،جامعة حلب

27) **الصباغ ، هبة علي طه** ،(2005). تحليل الإنحدار المضبب مع التطبيق. ماجستير علوم في الإحصاء، كلية علوم الحاسبات والرياضيات – جامعة الموصل

28) **الصباغ، هبة علي طه**؛الغنام،محمد طه أحمد. (2009). دراسة في المتغيرات المضببة والأنحدار المتعدد المضبب .مجلة تكريت للعلوم الادارية والاقتصادية.180-166 ,(14) .,

29) **نور الدين، محمد مالك،** (2017). تحسين التنبؤ باستخدام انحدار متجه الدعم عبر دراسة مسألة اختيار المعاملات. رسالة دكتوراه في الرياضيات–كلية العلوم–جامعة حلب.