استعمال الخوارزمية الجينية لتقدير معلمات توزيع كاما ليندلي

م.د.اياد حبيب شمال / جامعة ديالى / كلية الإدارة والاقتصاد / ayadstatistic@uodiyala.edu.iq
م.م.أرشد حميد حسن/ جامعة ديالى / كلية الإدارة والاقتصاد / Arshadeco@uodiyala.edu.iq

P: ISSN : 1813-6729
E : ISSN : 2707-1359
https://doi.org/10.31272/jae.i141.1007

تأريخ أستلام البحث : 2023/4/4
مقبول للنشر بتأريخ :2023/5/10

المستخلص

في هذا البحث تم عرض اهم التوزيعات الإحصائية الاحتمالية وهو توزيع كاما ليندل والناتج من دمج توزيع كاما مع توزيع ليندلي في حالة وجود معلمتين ويعتبر من التوزيعات المهمة في تطبيق الظواهر الاجتماعية والاقتصادية والطبيعية اذ تم دراسة اهم خواص توزيع كاما ليندلي مثل الدالة المولد للعزوم ودالة البقاء ودالة والمخاطرة .

ان دراسة اهم طرائق تقدير معلمات توزيع كاما ليندلي تتمثل بمقدرات الإمكان الأعظم ومقدرات الخوارزمية الجينية وبالاعتماد على بيانات حقيقية تمثل معدل سقوط الامطار لمدنية بعقوبة من عام 1990 - 2020 تم تقدير معلمات توزيع كاما ليندلي ، وكانت النتائج تدل على ملائمة البيانات لتوزيع كاما ليندلي وان مقدرات الخوارزمية الجينية أفضل من مقدرات الإمكان الأعظم من خلال الاعتماد على معيار حسن المطابقة اكيكي ومعيار حسن المطابقة بيزا اكيكي .

مجلة الادارة والاقتصاد
مجلد 48 العدد 141 / كانون الاول / 2023
الصفحات : 128 - 134

1. المقدمة introduction

تطبيقات الحياة الواقعية للتقنيات العددية المعاصرة في مجالات مختلفة مثل الطب والتمويل وعلوم الهندسة البيولوجية تلعب الإحصائيات دورًا مهمًا في هذه التطبيقات . غالبًا ما يتم استعمال التحليل الإحصائي الذي يعتمد بشدة على النماذج الاحصائية أو التوزيعات الاحتمالية المفترضة في التحليل، فإن العديد من المشاكل في الإحصاء لا تتبع أيًا من النماذج الاحصائية او التوزيعات الاحتمالية الكلاسيكية أو القياسية في التحليل . **(Nedjar & Zeghdoudi, 2016)**

لذا لجأ الباحثين الى استعمال توزيع جديد من مميزاته هو الدمج بين توزيعيين عن طريق دمج النسب للحصول على توزيع جديد بخصائص جديدة ومعلمات تقدير جديدة فان الكثير من الباحثين يهتمون في دمج توزيع لاندلي مع توزيعات أخرى للحصول على توزيع جديد ذات خصائص ومعلمات تقدير جديدة .

حيث درس الباحثين **Zamani & Ismail** عام 2010 توزيعا جديد مختلط عن طريق الدمج بين توزيع ذو الحدين السالب Negative Binomial Distribution بالمعالم (p , r) وتوزيع ليندلي Lindley distribution بالمعلمة θ فكان التوزيع المختلط الجديد يسمى التوزيع ذي الحدين السالب – ليندلي Negative Binomial-Lindley Distribution وتم تقدير معلمات التوزيع الجديد عن طريق مقدرات الإمكان الأعظم MLE ومقدرات العزوم ME وكانت النتائج تدل على ان التوزيع المختلط ذات مرونة وكفاءة عالية. ودرس الباحثين **Abdelmoezz & Mohamed** عام 2021 توزيع كوماراسوامي مع توزيع ليندلي من خلال توزيع مختلط يسمى توزيع كوماراسوامي ليندلي تم تقدير معلمات التوزيع المختلط باستعمال المحاكاة وكذلك تم تطبيق التوزيع الجديد على بيانات حقيقية تمثل بيانات البورصة المصرية من عام 2015 – 2019 . قدم الباحثين **Mahmoudi & Zakerzadeh عام** 2010 توزيع ليندلي بواسون المعمم Generalized Poisson–Lindley Distribution من خلال الدمج بين توزيع بواسون المركب وتوزيع ليندلي المعمم اذ تم مناقشة تقدير المعلمات للتوزيع المختلط من خلال مقدرات الإمكان الأعظم وتتم المقارنة مع توزيعات أخرى مشابه .

ودرس الباحثين **Zeghdoudi & Nedjar** عام 2016 توزيع ليندلي كاما Gamma Lindley distribution من خلال دراسة اهم الخصائص لتوزيع الجديد وكذلك تقدير معلماته من خلال مقدرات الإمكان الأعظم .

نلاحظ من خلال دمج توزيع ليندلي مع التوزيعات الاخرة ان من المهم في دراسة الخصائص الإحصائية لتوزيع المختلط هو تقدير معلماته المجهولة بالطرق الحديثة تختلف عن الطرق الكلاسيكية ، لذا سوف نقوم في هذا البحث تقدير معلمات توزيع كاما ليندلي من خلال الخوارزمية الجينية للحصول على مقدرات ذات كفاءة ومرونة عالية .

2. هدف البحث: Object of Research

يهدف البحث الى الاتي :

1- تقدير معلمات توزيع كاما ليندلي من خلال استعمال مقدرات الإمكان الأعظم واستعمال الخوارزمية الجينية لتقدير المعلمات .

2- المقارنة بين طرق تقدير المعلمات من خلال متوسط مربعات الخطأ Mean Square Error

3. توزيع كاما ليندلي: Gamma Lindley distribution(GLD)

تم تعريف توزيع ليندلي من قبل الباحث ليندلي **Lindley** عام 1958 وهو توزيع احادي المعلمة وتعرف دالة الكثافة الاحتمالية له كالاتي **(Zeghdoudi & Nedjar, 2016)** :

$$f(X,\theta)=\frac{\theta^2(1+X)e^{-\theta X}}{1+\theta} \qquad \theta, X>0 \qquad (1)$$

وان الدالة التراكمية لتوزيع ليندلي تكون وفق الصيغة الاتية:

$$F(X,\theta)=1-\frac{(\theta+1+\theta X)e^{-\theta X}}{1+\theta} \qquad (2)$$

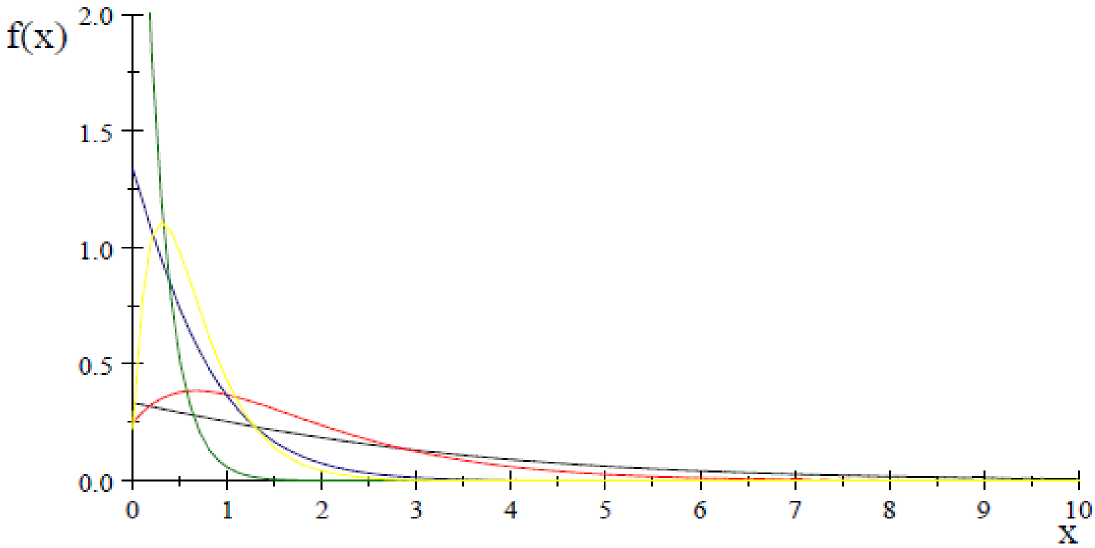
كما هو معرف ان توزيع ليندلي احادي المعلمة غير مرن لتحليل أنواع مختلفة من البيانات لذا تم اقتراح توزيع كاما ليندلي من قبل **Zeghdoudi & Nedjar.**

ليكن Y_1 يمثل متغير عشوائي يمتلك توزيع كاما $Y_1 \sim Gamma(2,\theta)$ و Y_2 يمثل متغير عشوائي يمتلك توزيع ليندلي $Y_2 \sim Lindley(\beta)$ وان المتغيرين العشوائيين مستقلين وليكن المتغير العشوائيX بشرط $X=Y_1, X=Y_2$ فأن دالة الكثافة الاحتمالية لتوزيع كاما ليندلي Gamma Lindley distribution تعطى بالصيغة الاتية :

$$f_{GaLd}(X,\theta,\beta)=\frac{\theta^2((\beta+\beta\theta-\theta)X+1)e^{-\theta X}}{\beta(1+\theta)} \qquad \beta\ ,\theta, X>0 \qquad (3)$$

توزيع كما ليندلي عندما $\beta = 1$ يصبح يمثل توزيع ليندلي، دالة التوزيع التراكمي لتوزيع كاما ليندلي تكون بالشكل الاتي:

$$F_{GaLd}(X) = 1 - \frac{((\theta\beta+\beta-\theta)(\theta X+1)+\theta)e^{-\theta X}}{\beta(1+\theta)} \quad \beta\ ,\theta, X > 0 \qquad (4)$$



الشكل رقم 1 يمثل دالة الكثافة الاحتمالية لتوزيع كاما ليندلي بمعلمات مختلفة.

4. بعض خواص توزيع كاما ليندلي Some properties of Gamma Lindley distribution

يمكن ملاحظة اهم الخواص الأساسية لتوزيع كاما ليندلي من خلال النقاط الاتية **(Nedjar & Zeghdoudi, 2016)** **(Zeghdoudi & Nedjar, 2016)**:

1- التوقع الى توزيع كاما لبندلي يكون وفق الصيغة الاتية:

$$\mu_X = E(X) = \frac{2\beta(1+\theta)-\theta}{\theta\beta(1+\theta)} \qquad ,E(X^2) = \frac{6\beta\theta+6\beta-4\theta}{\beta\theta^2+\beta\theta^3}$$

2- التباين الى توزيع كاما ليندلي يكون وفق الصيغة الاتية:

$$Var(X) = \frac{-(-2\beta\vartheta+\theta)^2+(2+6\theta)\beta^2-2\beta(\beta\theta-3\beta\theta^2+2\theta^2)}{\beta^2\theta^2(1+\theta)^2}$$

3- الدالة المولدة للعزوم لتوزيع كاما ليندلي تكون وفق الصيغة الاتية:

$$M_X(t) = e^{tX} = \left(\frac{\theta}{\theta-t}\right)^2 \left(\frac{\theta\beta+\beta-t}{\beta(\theta+1)}\right)$$

4- يمكن حساب المنول الى توزيع كاما ليندلي وفق الصيغة الاتية:

$$mod(X) = \begin{cases} \dfrac{\beta\theta+\beta-2\theta}{\theta(\beta+\beta\theta-\theta)^2} & \beta \in \{\frac{2\theta}{\theta+1}\ ,\infty\} \\ 0 & O.W \end{cases}$$

5- يمكن حساب دالة البقاء Survival Function ودالة معدل المخاطرة hazard rate function وفق الصيغ الاتية:

$$S_{GaLd}(X) = 1 - F_{GaLd}(X) = \frac{((\theta\beta+\beta-\theta)(\theta X+1)+\theta)e^{-\theta X}}{\beta(1+\theta)}$$

$$H_{GaLd}(X) = \frac{f(X,\theta)}{1-F(X,\theta)} = \frac{((\beta+\theta\beta-\theta)X+1)\theta^2}{\theta(\beta+\theta\beta-\theta)X+\beta+\theta\beta}$$

5. مقدرات الإمكان الأعظم: Maximum Likelihood Estimates (MLE)

ليكن X_i متغير عشوائي يمتلك توزيع كاما ليندلي فأن لن دالة الإمكان الأعظم يكون وفق الصيغة الاتية **(Zeghdoudi & Nedjar, 2016)**:

$$lnL(X_i, \theta, \beta) = 2nln\theta - nln\beta - nln(\theta+1) + \sum_{i=1}^{n} \ln\big((\beta+\theta\beta-\theta)X_i+1\big) - \theta\sum_{i=1}^{n} X_i \qquad (5)$$

وعند اشتقاق دالة الإمكان الأعظم بالنسبة للمعلمات θ , β يكون وفق الصيغ الاتية :

$$\frac{\partial nL(X_i,\theta,\beta)}{\partial\theta}=\frac{2n}{\theta}-\frac{n}{1+\theta}-\sum_{i=1}^{n}X_i+(\beta-1)\sum_{i=1}^{n}\frac{X_i}{(\beta+\theta\beta-\theta)X_i+1} \qquad (6)$$

$$\frac{\partial nL(X_i,\theta,\beta)}{\partial\beta}=\frac{-n}{\beta}+(1+\theta)\sum_{i=1}^{n}\frac{X_i}{(\beta+\theta\beta-\theta)X_i+1} \qquad (7)$$

ان المعادلات أعلاه لا يمكن إيجاد الحل لهم لذا سوف يتم استعمال طريقة تسجيل فشرfisher scoring وفق الصيغة الاتية:

$$\begin{bmatrix}\frac{\partial^2 nL(X_i,\theta,\beta)}{\partial\theta^2} & \frac{\partial nL(X_i,\theta,\beta)}{\partial\theta\partial\beta}\\ \frac{\partial nL(X_i,\theta,\beta)}{\partial\beta\partial\theta} & \frac{\partial^2 nL(X_i,\theta,\beta)}{\partial\beta^2}\end{bmatrix}_{\substack{\hat\theta=\theta_0\\ \hat\beta=\beta_0}}\begin{bmatrix}\hat\theta-\theta_0\\ \hat\beta-\beta_0\end{bmatrix}=\begin{bmatrix}\frac{\partial nL(X_i,\theta,\beta)}{\partial\theta}\\ \frac{\partial nL(X_i,\theta,\beta)}{\partial\beta}\end{bmatrix}_{\substack{\hat\theta=\theta_0\\ \hat\beta=\beta_0}} \qquad (8)$$

حيث ان

$$\frac{\partial^2 nL(X_i,\theta,\beta)}{\partial\theta^2}=\frac{-2n}{\theta^2}+\frac{n}{(1+\theta)^2}-(\beta-1)^2\sum_{i=1}^{n}\frac{{X_i}^2}{\left((\beta+\theta\beta-\theta)X_i+1\right)^2} \qquad (9)$$

$$\frac{\partial^2 nL(X_i,\theta,\beta)}{\partial\beta^2}=\frac{n}{\beta^2}-(1+\theta)^2\sum_{i=1}^{n}\frac{{X_i}^2}{(\beta+\theta\beta-\theta)X_i+1)^2} \qquad (10)$$

$$\frac{\partial nL(X_i,\theta,\beta)}{\partial\theta\partial\beta}=\frac{\partial nL(X_i,\theta,\beta)}{\partial\beta\partial\theta}\sum_{i=1}^{n}\frac{X_i}{(\beta+\theta\beta-\theta)X_i+1}-(\beta-1)(1+\theta)\sum_{i=1}^{n}\frac{{X_i}^2}{(\beta+\theta\beta-\theta)X_i+1)^2} \qquad (11)$$

6. الخوارزمية الجينية لتقدير معلمات توزيع كاما ليندلي: Genetic Algorithm Estimates (GAE)

يمكن استخدام الخوارزمية الوراثية (Genetic Algorithm) لتقدير معلمات توزيع كاما ليندلي (Gamma Lindley Distribution) . يتم ذلك عادةً من خلال الخطوات التالية**(Mitchell, 1998)** , **(O'Neill & others, 2009)**:

1- تعريف دالة الهدف :

يتم تعريف دالة الهدف بحيث تقوم بحساب قيمة دالة الاحتمال الشرطي لتوزيع كاما ليندلي باستخدام المعاملات المرشحة، ومن ثم حساب خطأ المعاملات عن القيم الفعلية للبيانات. يتم استخدام مقياس الخطأ (fitness function) كمعيار لقياس جودة المعاملات المرشحة.

2- تعريف المجال البحثي :

يتم تحديد المجال البحثي للمعاملات، مثل مجال القيم المسموح بها للمعاملات .

3- توليد الأفراد :

يتم إنشاء عدد معين من الأفراد (individuals) الذين يحتوي كل منهم على معاملات مختلفة في المجال البحثي ويتم إنشاء الأفراد بشكل عشوائي وفقاً للمجال البحثي .

4- عملية التطوير :

تتم عملية التطوير (evolution process) بحيث تتمثل في العمليات التالية:

- **عملية التقييم :** يتم حساب قيمة دالة الهدف لكل فرد .
- **عملية التنافس :** يتم اختيار الأفراد الأفضل من بين الأفراد المتاحين وتركهم للأجيال القادمة .
- **عملية التكاثر :** يتم تكاثر الأفراد المختارين باستخدام المعاملات الخاصة بهم لإنتاج أفراد جدد .
- عملية الطفرة: يتم إجراء عمليات الطفرة لبعض الأفراد المتكاثرين لزيادة التنوع في الأجيال القادمة .

5- الحد من عدد الأجيال :

تتكرر عملية التطوير حتى تصل الأفراد إلى معاملات يمكن القول إنها تقدم أفضل قيم ويتم تحديد عدد الأجيال المطلوبة قبل بدء العملية، ويمكن زيادة عدد الأجيال في حال لم يتم الحصول على أفضل النتائج الممكنة .

6- الاختبار والتحليل :

يتم اختبار المعاملات النهائية على مجموعة بيانات مختلفة لتقييم كفاءة الخوارزمية. كما يتم تحليل النتائج المتحصل عليها وتقييم جودة الخوارزمية .

يمكن استخدام الخوارزمية الوراثية لتقدير المعلمات المختلفة لتوزيع كاما ليندلي، مثل المعلمات المتعلقة بالشكل والمعلمات المتعلقة بالمقياس. يجب أن يتم تحديد المجال البحثي والدالة الهدف بعناية، ويمكن تعديل خطوات الخوارزمية وفقًا لمتطلبات المشكلة المعينة.

7. معيار حسن المطابقة Goodness of Fit

لمعرفة هل ان البيانات تلائم التوزيع المناسب لها في التحليل الاحصائي من المواضيع المهمة اذ يوجد العديد من المعايير الإحصائية لمعرفة هل البيانات التي تم استعمالها تتبع توزيع احتمالي معين وكلي نحصل على معلومات دقيقة حول البيانات المستخدمة في هذا البحث سوف يتم استعمال معياريين لحسن المطابقة وهما كالاتي **(Tahir , 2015), (Bozdogan, 2000)**:

1- معيار اكيكي : Akaike's Information Criterion (AIC)

ان الصيغة العامة لهذا المعيار تكون كالاتي:

$$AIC = -2logL + 2K \qquad (12)$$

حيث ان $logL$ دالة لوغارتم الامكان لتوزيع كاما ليندلي.

K عدد المعلمات لدالة التوزيع.

2- معيار بيز اكيكي: Bayesian Information Criterion (BIC)

ان الصيغة العامة لهذا المعيار تكون كالاتي:

$$BIC = -2logL + K\log(n) \qquad (12)$$

حيث ان $logL$ دالة لوغارتم الامكان لتوزيع كاما ليندلي.

K عدد المعلمات لدالة التوزيع و **n** عدد المشاهدات

8. الجانب التطبيقي :

يعد توزيع كاما ليندلي من التوزيعات الإحصائية التي يتم استعمالها في الكثير من الظواهر الاجتماعية والاقتصادية والطبيعية لذا سوف يتم تطبيق دالة التوزيع على بيانات طبيعية تمثل معدل سرعة الرياح لمدينة بعقوبة خلال السنوات من 1990-2020 وكانت البيانات متمثلة من خلال الجدول الاتي :

الجدول رقم 1 يبين معدل سرعة الرياح لمدينة بعقوبة

السنوات	معدل سرعة الرياح كم/ساعة	السنوات	معدل سرعة الرياح كم/ساعة
1990	3.48	2006	3.49
1991	3.46	2007	3.4
1992	3.48	2008	3.35
1993	3.47	2009	3.36
1994	3.47	2010	3.39
1995	3.44	2011	3.5
1996	3.43	2012	3.39
1997	3.46	2013	3.45
1998	3.33	2014	3.41
1999	3.53	2015	3.34
2000	3.52	2016	3.36
2001	3.52	2017	3.4
2002	3.42	2018	3.35
2003	3.49	2019	3.33
2004	3.41	2020	3.37
2005	3.41		

من خلال البيانات في الجدول أعلاه التي تمثل معدل سرعة الرياح لمدينة بعقوبة يتم تقدير معلمات توزيع كاما ليندلي $\hat{\theta}$ و $\hat{\beta}$ والجدول الاتي يبين تقدير المعلمات:

الجدول رقم 2 يبين تقدير المعلمات لتوزيع كاما ليندلي

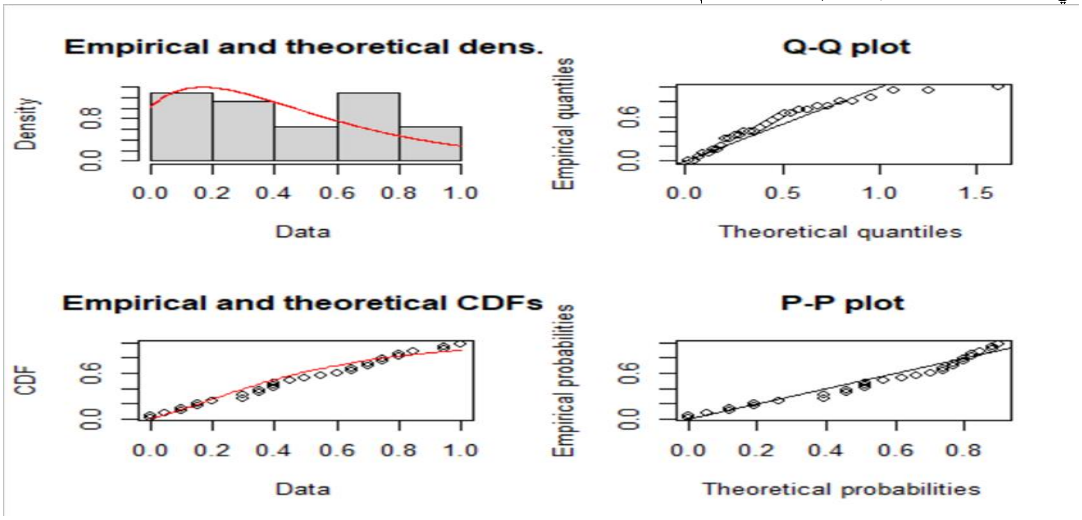
parameters estimation	MLE	GAE
$\hat{\theta}$	3.563889	1.006659
$\hat{\beta}$	2.719026	0.1296338

الجدول أعلاه يبين تقدير المعلمات لتوزيع كاما ليندلي باستعمال مقدرات الإمكان الأعظم MLE ومقدرات الخوارزمية الجينية GAE حيث كانت المقدرات بطريقة الخوازمية الجينية اقل من المقدرات بطريقة الإمكان الأعظم.

الجدول رقم 3 يبين معيار حسن المطابقة

Goodness of Fit	AIC	BIC
MLE	16.52953	19.3975
GAE	5.576369	15.39116

الجدول أعلاه يبين معيار معلومات اكيكي ومعيار معلومات بيز بالاعتماد على مقدرات الإمكان الأعظم ومقدرات الخوارزمية الجنية، نجد من خلال المعياريين أعلاه ان مقدرات االخوارزمية الجينية كانت أفضل من مقدرات الإمكان الأعظم في تقدير معلمات توزيع كاما ليندلي حيث كانت نتائج المعياريين باستعمال مقدرات الخوارزمية الجينية اقل من في حالة استعمال مقدرات الإمكان الأعظم.



الشكل (2) يمثل دالة الكثافة الاحتمالية ودالة التوزيع التجمعية

من الشكل أعلاه نلاحظ ملائمة تقديرات معلمات توزيع كاما ليندلي في حالة استعمال الخوارزمية الجينية في التقدير لبيانات معدل سرعة الرياح لمدينة بعقوبة، كما يبين الشكل Q-Q plot والشكل p-p plot ان منحى التقدير متقارب مع نقاط الانتشار.

9. الاستنتاجات

يمكن تلخيص اهم الاستنتاجات التي تم التوصل اليها من خلال هذا البحث من خلال النقاط الاتية

1- من خلال مقياس حسن المطابقة لمعيار AIC ومعيار BIC مقدرات الخوارزمية الجينية كانت أفضل من مقدرات الإمكان الأعظم لتقدير معلمات توزيع كاما ليندلي.

2- معلومات حسن المطابقة الخاصة بمعيار BIC كانت أفضل من معيار AIC.

3- من خلال معاير حسن المطابقة البيانات الخاصة بمعدل سقوط الامطار لمدينة بعقوبة كان التوزيع الاحتمالي المناسب لها هو توزيع كاما ليندلي.

المصادر:

1- Abdelmoezz, S., & Mohamed, S. M. (2021). The Kumaraswamy Lindley Regression Model with Application on the Egyptian Stock Exchange: Numerical study, Regression model. Jurnal Matematika, Statistika dan Komputasi, 18(1), 1-11.

2- Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. Journal of mathematical psychology, 44(1), 62-91.

3- Mahmoudi, E., & Zakerzadeh, H. (2010). Generalized poisson–lindley distribution. Communications in Statistics—theory and Methods, 39(10), 1785-1798.

4- Mitchell, M. (1998). An introduction to genetic algorithms. MIT press.

5- Nedjar, S., & Zeghdoudi, H. (2016). On gamma Lindley distribution: Properties and simulations. Journal of Computational and Applied Mathematics, 298, 167-174.

6- O'Neill, M. (2009). Riccardo Poli, William B. Langdon, Nicholas F. McPhee: A Field Guide to Genetic Programming: Lulu. com, 2008, 250 pp, ISBN 978-1-4092-0073-4.

7- Tahir ,M.H. (2015), " The Weibull Lomax Distribution Probabilities and Applications" Journal of Mathematics and statistics . Vol. 44,pp.461-480.

8- Zamani, H., & Ismail, N. (2010). Negative binomial-Lindley distribution and its application. Journal of mathematics and statistics, 6(1), 4-9.

9- Zeghdoudi, H., & Nedjar, S. (2016). Gamma Lindley distribution and its application. Journal of Applied Probability and Statistics, 11(1), 129-138.

Using genetic algorithm to estimate gamma Lindley distribution parameters

Iyad Habib Shamal / University of Diyala / College of Administration and Economics / ayadstatistic@uodiyala.edu.iq
Arshad Hamid Hassan / University of Diyala / College of Administration and Economics / Arshadeco@uodiyala.edu.iq

Abstract

In this research, the most important statistical probability distributions were presented, which is the Kamma Lindley distribution, which results from combining the Kamma distribution with the Lindley distribution in the case of the presence of two parameters. It is considered one of the important distributions in the application of social, economic, and natural phenomena, as the most important properties of the Kamma Lindley distribution were studied, such as the moment generating function, the survival function, and the survival function. And risk.
The study of the most important methods for estimating the parameters of the Kamma Lindley distribution is represented by the maximum potential estimators and the estimators of the genetic algorithm, and based on real data representing the rainfall rate for the city of Baqubah from the year 1990 - 2020, the parameters of the Kamma Lindley distribution were estimated, and the results indicated the suitability of the data to the Kamma Lindley distribution and the algorithm estimators. Genetic is better than the maximum possibility estimator by relying on the Akiki goodness of fit criterion and the Pisa goodness of fit criterion.

****
****
****