



استعمال الخوارزمية الجينية في تقدير معلمات دالة التوزيع الاسي الفوقي بواسطة المحاكاة

سلام جاسم محمد
جامعة الكرخ للعلوم, شارع حيفا, بغداد,
العراق

salam.j@kus.edu.iq

Received: 20/9/2020

Accepted :18/10/2020

Published :FEBRUARY / 2021

هذا العمل مرخص تحت اتفاقية المشاع الابداعي نسب المُصنَّف - غير تجاري - الترخيص العمومي الدولي 4.0

[Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)



مستخلص البحث

في هذا البحث تم التركيز على تقدير معلمات دالة التوزيع الاسي الفوقي بأستعمال طريقة الامكان الاعظم و الخوارزمية الجينية وقد تم اعتماد اكثر من معيار ومنها MSE للمقارنة وذلك بأستعمال اسلوب المحاكاة.

المصطلحات الرئيسية للبحث: متوسط معدل الوصول, دالة كثافة الاحتمال, دالة التوزيع التراكمي, التوزيع الأسّي الفوقي, الخوارزمية الجينية, دالة الامكان الاعظم.

المقدمة :

تعد عملية تقدير المعلمات جزء مهم في موضوع الاستدلال الاحصائي اضافة الى اختبار الفرضيات , ويتم التقدير بواسطة جمع المعلومات عن معلمة المجتمع وذلك بالاعتماد على النتائج المستخرجة من العينة المسحوبة من المجتمع , علماً ان هناك طرق مختلفة لعملية التقدير والهدف الاساسي هو الحصول على مقدرات ذات خصائص جيدة , خاصة في حالة وجود اكثر من طريقة لتقدير المعلمات وان ذلك سيؤدي الى المفاضلة بين هذه المقدرات لأختيار الافضل منها.

ان هدف البحث هو تقدير معلمات دالة التوزيع الاسي الفوقي بأستعمال طريقة الامكان الاعظم و الخوارزمية الجينية وقد تم اعتماد اربعة معايير لمعرفة أفضل نموذج موضحة في الجانب التجريبي وفي جدول رقم (1) بحجوم عينات مختلفة , وكذلك تم الاعتماد على صيغة متوسط مربعات الخطأ (MSE) لمعرفة أفضل نموذج موضحة في الجانب التجريبي وفي جدول رقم (2) بحجوم عينات مختلفة وذلك بأسلوب المحاكاة , علماً الغرض من البحث هو و تقدير معلمات دالة التوزيع الاسي الفوقي بأستعمال الخوارزمية الجينية ومقارنتها مع طريقة الامكان الاعظم بأسلوب المحاكاة .

1. التوزيع الاسي exponential distribution [1] [2]

التوزيع الاسي هو احد التوزيعات المهمة وخاصة في مجال احتساب دالة المعولية والمجالات الهندسية , ان النموذج الاسي الأكثر شيوعاً هو النموذج الاسي ذو معلمة واحدة θ وتمثل معدل او متوسط الحياة او متوسط الوقت المستغرق لحين الفشل (MTTF) وهي تسلك سلوك معلمة القياس بينما تمثل المعلمة $\lambda = \frac{1}{\theta}$ متوسط معدل الوصول (MAR) Mean arrival rate

اما الدالة الاحتمالية (PDF) لهذا التوزيع فهي:

$$f(t) = \lambda e^{-\lambda t}, t \geq 0 \dots (1)$$

$$f(t) = \frac{1}{\theta} e^{-\left(\frac{t}{\theta}\right)}, \theta \geq 0 \dots (2)$$

ان الدالة الاحتمالية التجميعية (CDF) لهذا التوزيع هي:

$$\begin{aligned} F(t) &= p_r(T \leq t) \\ &= \int_0^t f(u) du \\ &= 1 - p_r(T > t) \\ F(t) &= 1 - e^{-\lambda t} \dots (3) \end{aligned}$$

2. التوزيع الاسي الفوقي Hypoexponential distribution [3] [5]

وهو عبارة توزيع مجموع $2 \leq$ من المتغيرات العشوائية الأسية المستقلة يستخدم هذا التوزيع في نمذجة مراحل الأسية المتعددة في السلسلة .

افرض ان x_1, x_2, \dots, x_n متغيرات مستقلة معلمة $\alpha_i > 0$ ستكون دالة التوزيع الاسي الفوقي (pdf) كالآتي :

$$f x_i(x) = \alpha_i e^{-\alpha_i x}, x > 0 \dots \dots (4)$$

وعليه سيكون المتغير العشوائي $s = \sum_{i=1}^m X_i$ يتبع التوزيع الاسي الفوقي بمعلمة على شكل متجة $a = (\alpha_1, \dots, \alpha_m)$, وستكون الدالة بالشكل الآتي:

$$f_s(x) = \sum_{i=1}^m \frac{1}{p_i} e^{-\alpha_i x}, x > 0 \dots \dots (5)$$

حيث ان :

$$p_i = \prod_{j=1}^m \left(1 - \frac{\alpha_j}{\alpha_j}\right)$$

الدالة التجميعية التراكمية (CDF) ستكون بالشكل الآتي :

$$F_s(x) = 1 - \sum_{i=1}^m \frac{1}{p_i} e^{-\alpha_i x} \dots \dots (6)$$

ويمكن تعريف التوزيع الآسي الفوقي الذي يرمز له بالرمز $EH(a, \lambda)$ ، بالدالة الاحتمالية التالية:

$$f_y(x) = \lambda e^{-\lambda x} \sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda x})^{\alpha_i - 1}, x > 0 \dots \dots (7)$$

وعليه ستكون الدالة الاحتمالية التجميعية (CDF) هي:

$$F_y(x) = \sum_{i=1}^m \frac{1}{p_i} (1 - e^{-\lambda x})^{\alpha_i}, x > 0 \dots \dots (8)$$

3. طرائق التقدير

3.1 طريقة الامكان الاعظم Maximum likelihood estimation [3]

وهي احدى اهم طرائق التقدير التي تهدف الى جعل دالة الامكان في نهايتها العظمى وسيتم التقدير الى التوزيع الآسي الفوقي , فاذا كانت لدينا عينة عشوائية (y_1, y_2, \dots, y_n) تتوزع توزيع الآسي الفوقي، فإن مقدر الامكان الاعظم هو الذي يجعل دالة الامكان في نهايتها العظمى ويمكن الحصول عليها باشتقاق لوغاريتم دالة الامكان ومساواتها بالصفر، فاذا كانت (x) تتوزع التوزيع الآسي الفوقي ستكون كالآتي :-

$$f_y(x) = \lambda e^{-\lambda x} \sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda x})^{\alpha_i - 1}, x > 0 \dots \dots (9)$$

$$L(\theta) = \lambda^n e^{-\lambda \sum_{u=1}^n y_u} \prod_{y=1}^n \left(\sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 1} \right) \dots \dots (10)$$

ستكون الدالة اللوغارتمية كالآتي:

$$\ell(\theta) = \log(L(\theta)) = n \log(\lambda) - \lambda \sum_{u=1}^n y_u + \left(\sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 1} \right) \dots \dots (11)$$

اشتقاق المعادلة (11) ومساواتها بالصفر سنحصل على :

$$\frac{\partial \ell(\theta)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{u=1}^n y_u + \sum_{u=1}^n y_u e^{-\lambda y_u} \frac{\sum_{i=1}^m \frac{\alpha_i (\alpha_i - 1)}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 2}}{\sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 1}} = 0 \dots \dots (12)$$

ولأي $q \in (1, \dots, m)$

$$\frac{\partial \ell(\theta)}{\partial \alpha_q} = \sum_{u=1}^n \frac{\frac{\partial}{\partial \alpha_q} \left(\sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 1} \right)}{\sum_{i=1}^m \frac{\alpha_i}{p_i} (1 - e^{-\lambda y_u})^{\alpha_i - 1}} = 0 \dots \dots (13)$$

هذه المعادلات ليست قابلة للحل من الناحية التحليلية ، وعليه سيتم تقدير المعلمات من خلال طريقة (نيوتن رافسن) (Newton-Raphson) [4] وهي خوارزمية فعالة لإيجاد جذور تابع حقيقي. لذلك تعتبر مثالا لخوارزميات إيجاد الجذور. يمكن استعمالها لإيجاد الحدود العليا والحدود الدنيا لمثل هذه التوابع، عن طريق إيجاد جذور المشتق الأول للتابع

3.2 الخوارزمية الجينية (Genetic Algorithm) [6] [7]

تعتبر الخوارزمية الجينية GA واحدة من طرق البحث المبنية على الية الانتقاء (الاختيار) الطبيعي وعلم الوراثة الطبيعي. تصنف الخوارزمية الجينية كواحدة من الخوارزميات التطورية (Evolutionary Algorithm) والمبنية على اساس محاكاة عمل الطبيعة من منظور العالم دارون. تعمل هذه الخوارزمية كطريقة بحث عشوائي لغرض ايجاد حلول مثلى او قريبة من المثلى عن طريق تحقيق مبدأ الامثلية واستعمال اليات احيائية طبيعية مثل الوراثة والتزاوج والطفرة الوراثة وتعتبر هذه الخوارزمية من التقنيات الحديثة الهامة في مجال البحث عن الحل الامثل من بين مجموعة من الحلول المتوفرة من خلال تمرير الصفات الجيدة لعمليات التوليد المتعاقب وانتاج ذرية امثل وتكرار الدورات الوراثة لتحسين الذرية باطوار وانماط حديثة.

خطوات الخوارزمية الجينية

تتألف الخوارزمية الجينية من الخطوات التالية:

1- التهيئة (Inialization): وهي الخطوة الاولى في الخوارزمية الجينية حيث يتم توليد مجموعة حلول عشوائية على شكل كروموسومات وان طول الكروموسوم وطريقة تمثيله يعتمد على طبيعة المشكلة وتوجد هنالك اربع طرق اساسية لتمثيل وترميز الكروموسومات بشكل يسهل التعامل معها من قبل الحاسوب اعتماداً على طبيعة المشكلة المراد حلها. وعادة يوجد هناك عدة منات او الاف من الحلول الممكنة بشكل تقليدي يتم توليد الكروموسومات بشكل عشوائي بحيث تعطي مجموعة كاملة من الحلول الممكنة للبحث الفضائي (Search Space) , ومن طرق تمثيل الكروموسومات التمثيل او الترميز الثنائي (Binary Encoding) في هذا النوع يتم تمثيل الكروموسومات على شكل سلسلة من الارقام التي تضم الصفر والواحد فقط لاغير وكما هو موضح في ادناه:

| | |
|-----------------------|------------|
| 1 0 0 1 0 1 1 1 0 0 1 | كروموسوم 1 |
| 0 1 0 0 1 0 1 0 0 1 1 | كروموسوم 2 |

2- الاختيار (Selection) هذه العملية تطبق على جميع الاجيال المتعاقبة اذ يتم اختيار مجموعة من الكروموسومات وفق نسبة معينة لغرض انتاج وتوليد جيل جديد. عملية الاختيار هذه تعتمد ايضا على قيمة دالة المفاضلة (Fitness Function). هنالك طريقة اخرى للاختيار تكون عن طريق اختيار مجموعة عشوائية من الكروموسومات ومن مساوي هذه الطريقة هو احتياجها لوقت طويل جداً. وتعتبر طريقة عجلة الروليت من اكثر الطرق استخداماً في الاختيار.

3- اعادة الانتاج (Reproduction) وهي عملية توليد وانتاج جيل جديد من الافراد التي تم انتقائها واختيارها من خلال عملية الانتقاء ثم عملية التزاوج (Crossover) والطفرة الوراثة (Mutation) لانتاج الابناء لتشكيل الجيل الجديد وكما يأتي.

أ. عملية التزاوج Crossover

يتم اجراء عملية التزاوج على الابوين الذين تم اختيارهم عن طريقة عملية الاختيار او الانتقاء لغرض توليد فردين جديدين وتستمر هذه العملية لغاية تكوين الجيل الجديد بكافة افراده المتفق عليهم. توجد عدة انواع من انواع التزاوج، منها ما يأتي:

➤ التزاوج من خلال نقطة تزاوج واحدة (نقطة قطع واحدة) One point crossover

➤ التزاوج من خلال اكثر من نقطة تزاوج (نقطتي قطع او اكثر) K-Points Crossover

ب. عملية الطفرة الوراثة Mutation

وهي عملية تغيير مفاجئ في الابناء المتولده من خلال عملية التزاوج بحيث تؤدي الى تغيير في شكل الكروموسوم عن طريق تغيير احدى جينات الكروموسوم (تغيير بت واحد او اكثر) وهذه العملية ليست ناتجة عن الاباء حيث ان عملية اعادة التوليد او الانتاج تؤدي الى توليد وانتاج الكروموسومات الجديدة التي يتم تطبيق دالة المفاضلة عليها لغرض حساب قيمة المفاضلة التي تساعد في عملية انتاج وتوليد الجيل الجديد. هنالك عدة انواع من الطفرة الوراثة نستعرض فيما يلي بعضاً منها

➤ التغيير مع التزحيف Shaft and Exchange

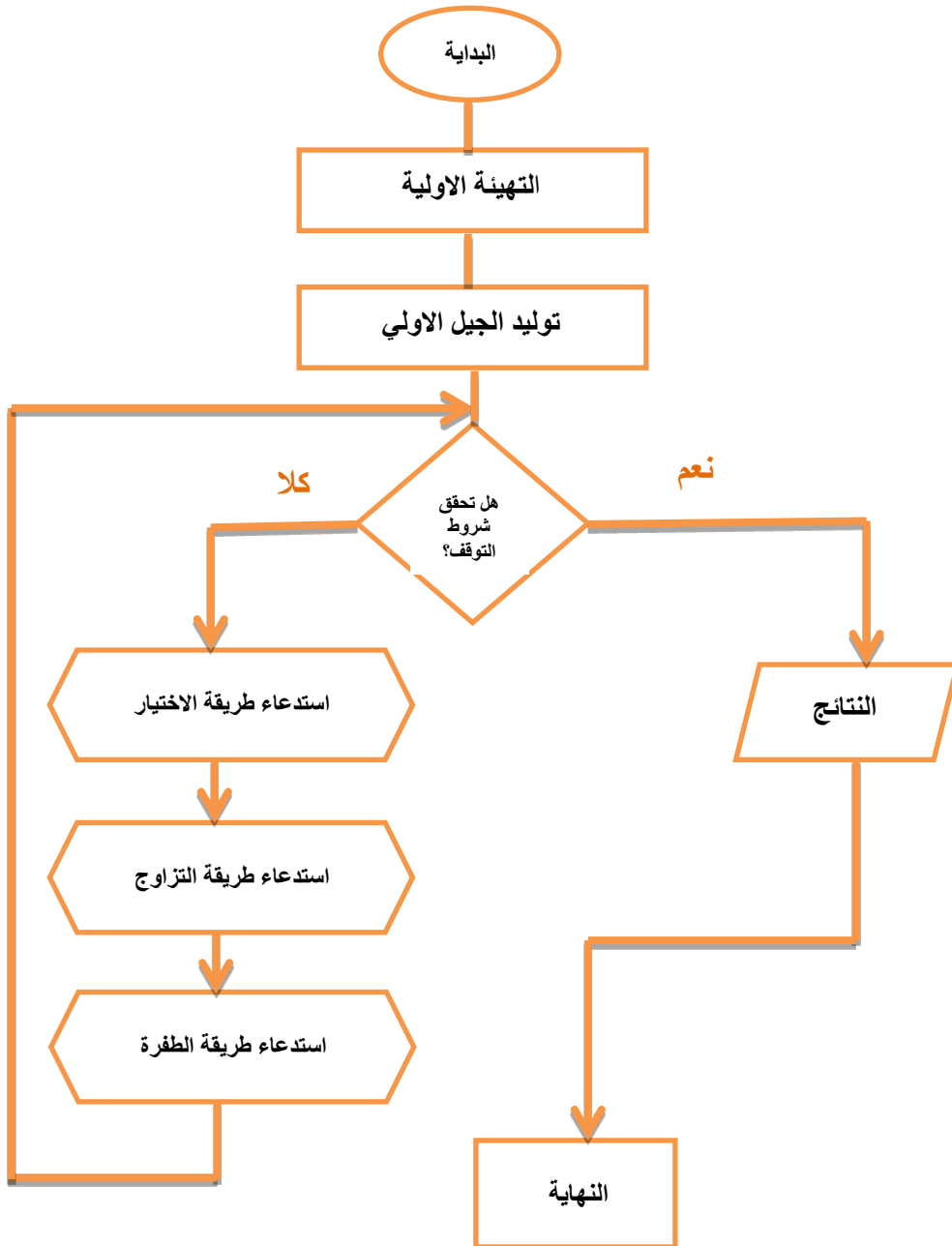
➤ التغيير فقط Exchange

4- الانتهاء والتوقف Termination Conditions

تنتهي وتتوقف الخوارزمية الجينية عند حدوث احد الاسباب والعوامل التالية:

- ايجاد الحل الامثل.
- الوصول الى عدد الاجيال المطلوبة.
- الوصول الى قيمة معينة مثل كلفة الانتاج.
- الوقوع في ال Local Minimum وعدم المقدرة على الخروج منها.

الشكل التالي يوضح خطوات الخوارزمية الجينية



المخطط الانسيابي للخوارزمية الجينية

الجانب التجريبي

ان اسلوب المحاكاة يعتبر من افضل الاساليب في حل المشاكل المعقدة التي لا يمكن حلها في الواقع الخارجي وكذلك يتم الرجوع الى هذا الاسلوب في حالة عدم توفر البيانات الاحصائية الملائمة , علماً ان هذا الاسلوب يوفر كثيراً من الجهد والمال ويتم ذلك من خلال الخطوات الآتية :

1. تحديد حجوم العينات : تم اختيار اربع حجوم للعينات هي (150,75,50,20)
2. توليد ارقام عشوائية : تم توليد ارقام عشوائية U_i تتبع التوزيع المنتظم ضمن الفترة (0,t) .

$$U_i \sim U(0,1) \quad , i = 1, 2, \dots, n$$

U_i : يمثل متغير عشوائي مستمر يتم توليده باستخدام الحاسبة الالكترونية على وفق الصيغة الآتية :-

$$U = RND \dots\dots(14)$$

3. تحويل البيانات التي تم توليدها في المعادلة (14) والتي تتبع التوزيع المنتظم $U(0,1)$ الى بيانات تتبع التوزيع الاسي الفوقي من خلال استخدام دالة التوزيع التجميعية (cdf) وكالاتي:

$$F_y(x) = \sum_{i=1}^m \frac{1}{p_i} (1 - e^{-\lambda x})^{\alpha_i} \quad \dots\dots(15)$$

ومن ثم فإن :

$$U = 1 - \sum_{i=1}^m \frac{1}{p_i} (1 - e^{-\lambda x})^{\alpha_i} \quad \dots\dots(16)$$

مقياس المقارنة [3]

أ. تم الاعتماد على اربعة معايير في (جدول رقم 1) ويمكن تعريفها كالاتي :

- $-2\log(L)$: (L) هو لوغاريتم دالة الامكان

- AIC : (Akaike Information Criterion) (مقياس معلومات اكاكي)

ويعرف بكونه مقياس للجودة النسبية لنموذج احصائية ما , المطبقة على مجموعة بيانات مقياس اكاكي مستلهم من نظرية المعلومات ويساعد على اختيار النموذج الافضل وصيغته

$$AIC = -2 \log(L) + 2k$$

- AICC : (Akaike Information Criterion Corrected) (تصحيح مقياس معلومات اكاكي)

يظهر AIC درجة عالية محتملة من التحيز السلبي في تطبيقات العينات الصغيرة يعرف AICC انه مصحح لهذا التحيز

وصيغته:

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}$$

- BIC : (Bayesian Information Criterion) (مقياس معلومات بيز) :

يعرف بكونه مقياس لاختيار النموذج بين مجموعة محدودة من النماذج , يفضل استعمال النموذج ذو أقل

(BIC) وهي تستند جزئياً الى دالة الاحتمالية وترتبط ارتباطاً وثيقاً بمقياس معلومات (AIC)

وصيغته :

$$BIC = -2 \log(L) + k \log(n)$$

حيث يشير k إلى عدد المعلمات المقدرة و n إلى حجم العينة.

أفضل تقدير عندما تكون القيم المستخلصة من الصيغ اعلاه اقل ما يمكن.

جدول رقم (1)

| Sample size | Methods | -2log(L) | AIC | AICC | BIC |
|-------------|----------|-----------|-----------|-----------|-----------|
| 20 | MLE(N.R) | 60.93266 | 64.93266 | 65.63854 | 66.92412 |
| | Genetic | -4.31818 | -4.31778 | -4.31771 | -4.31758 |
| 50 | MLE(N.R) | 75.16744 | 79.16744 | 79.42276 | 82.99149 |
| | Genetic | 335.42088 | 339.42088 | 339.67620 | 343.24492 |
| 75 | MLE(N.R) | 77.44134 | 81.44134 | 81.60801 | 86.07632 |
| | Genetic | 629.50500 | 633.50500 | 633.67166 | 638.13997 |
| 150 | MLE(N.R) | 38.31409 | 42.31409 | 42.39573 | 48.33536 |
| | Genetic | 578.24761 | 582.24761 | 582.32924 | 588.26888 |

حيث ان :

MLE(N.R): يقصد بها التقدير بطريقة الامكان الاعظم بأستعمال نيوتن رافسون.

Genetic: يقصد بها التقدير بأستعمال الخوارزمية الجينية.

ب. تم الاعتماد ايضا على صيغة متوسط مربعات الخطأ لمعرفة أفضل نموذج في (جدول رقم 2) الاتية:

$$MSE(\hat{S}) = \frac{\sum_{i=1}^R (y_i - \hat{y})^2}{R}$$

إذ أن :

 \hat{y} : النموذج المقدر

R: عدد التكرارات Replication لكل تجربة والذي كان مساويا الى 100.

تم اعتماد المحاكاة لتقدير المعلمة λ والمعلمة α للتوزيع الآسي الفوقي باستخدام برنامج (matlab)

جدول رقم (2)

| Sample size | Methods | Lmda | Alfa | MSE | Best |
|-------------|----------|---------|---------|-------------------|----------|
| 20 | MLE(N.R) | 0.12193 | 0.03458 | 0.089638733738640 | Genetic |
| | Genetic | 0.00011 | 0.06174 | 0.000000042472502 | |
| 50 | MLE(N.R) | 0.47424 | 0.07436 | 0.050513574522692 | Genetic |
| | Genetic | 0.03633 | 4.99271 | 0.016567707923437 | |
| 75 | MLE(N.R) | 0.47881 | 0.11722 | 0.023039181658704 | Genetic |
| | Genetic | 0.01527 | 8.84103 | 0.016427905123298 | |
| 150 | MLE(N.R) | 0.86075 | 0.23653 | 0.008647425285253 | MLE(N.R) |
| | Genetic | 0.00168 | 1.13123 | 0.020119605017839 | |

الاستنتاجات

1. اظهرت النتائج في جدول رقم (1) انه في حجم العينة (20) تكون طريقة الخوارزمية الجينية افضل في كافة المعايير الاربعة , اما في حجم العينة (150,75,50) فان طريقة الامكان الاعظم افضل للمعايير الاربعة.
2. بينت النتائج بأن في جدول رقم (2)، ان مجموع مربعات الخطأ MSE للنموذج يكون افضل اي اقل في طريقة الخوارزمية الجينية في حجومات العينات (75,50,20) ماعدا العينة ذات الحجم (150) تكون طريقة الامكان الاعظم هي الافضل.

التوصيات

1. يوصي الباحث بأعتماد طريقة الامكان الاعظم في حالة العينات الكبيرة للمعايير الاربعة في جدول رقم (1) وكذلك لمتوسط مربعات الخطأ في جدول رقم (2).
2. يوصي الباحث بأعتماد طريقة الخوارزمية الجينية للعينات الصغيرة للمعايير في جدول رقم (1) وكذلك لمتوسط مربعات الخطأ في جدول رقم (2).
3. يوصي الباحث بتوسع نطاق الدراسة وذلك بأستعمال طرق تقدير اخرى.

References

1. Abdul Ahad, Attaf. Edward (2007). Reliability estimates of the two-parameter exponential distribution, a comparative study. Master Thesis submitted to the College of Administration and Economics, University of Baghdad, 2 (11).
- 2- Cohen , A.c and Helm , F.R. (1973) . Estimation in Exponential Distribution Technometrics. American Statistical Association and American Society for Quality Stable. 15(2) , 415-418.
- 3- Christophe C. (2017) .A new family of distributions based on the hypoexponential distribution with fitting reliability data, University of Caen Basse-Normandie,(2)
- 4- Endre Suli and David F. Mayer(2003) . An Introduction to Numerical Analysis. Cambridge University Press.
- 5- Khaled Smaili, Therrar Kadri and Seifedine Kadry , (2013). Hypoexponential Distribution with Different Parameters. Applied Mathematics, 4, 624-631
<http://dx.doi.org/10.4236/am.2013.44087> Published Online
(<http://www.scirp.org/journal/am>)
- 6- Mitchell , M .(1998).An Introduction To Genetic Algorithms. the United States: Cambridge University Press.
7. Sabah Manfi Rida and Yasser Ibrahim Saleh,(2015). The use of the genetic algorithm in solving the problem of multi-target allocation, Journal of the Iraqi Society for Information Technology Issue 723.

The use of the genetic algorithm to estimate the parameters function of the hypoexponential distribution by simulation

Salam Jasim Mohammed
Al-Karkh University of Science, Haifa
Street Baghdad, Iraq
salam.j@kus.edu.iq

Received: 20/9/2020

Accepted :18/10/2020

Published :FEBRUARY / 2021



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

ABSTRACT

In this research, focus was placed on estimating the parameters of Hypoexponential distribution function using the maximum likelihood method and genetic algorithm. More than one standard, including MSE, has been adopted for comparison by Using the simulation method..

Keywords: Mean arrival rate; probability density function; cumulative distribution

Function; Hypoexponential distribution; Genetic algorithm; Maximum likelihood estimation.