AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

مقارنة بعض الطرائق لتقدير دالة المخاطرة لأنموذج Gompertz المعمم باستعمال المحاكاة

م.م. منال محمود رشيد

manal4254.@gmail.com

المديرية العامة لتربية بغداد الرصافة الثالثة، بغداد، العراق

المستخلص

تم في هذا البحث دراسة أحد أهم نماذج الفشل الواسعة الاستعمال في دراسة المعولية واختبارات الحياة عندما يكون المجتمع غير متجانس وهو توزيع Gompertz المعمم إذ تم تناول هذا الأنموذج بشيء من التفصيل من حيث أهميته واستعمالاته وصياغته، ومن ثم التركيز على بعض طرائق تقدير دالة االمخاطرة حيث تم استعمال اثنتين من الطرائق المهمة وهما طريقة الجينية وخوارزمية Nelder Mead والمقارنة بين هذه الطرائق تم استعمال جانب المحاكاة بطريقة مونت كارلو حيث يتميز هذا الاسلوب بالمرونة ويوفر الكثير من التكاليف عن طريق الأخذ بنظر الاعتبار حجوم العينات المختلفة (كبيرة ومتوسطة وصغيرة) والقيم الافتراضية المختلفة لمعلمات التوزيع وتكرار التجربة في كل مرة ويتم في هذا الاسلوب توليد البيانات دون اللجوء الى البيانات الحقيقية واثبتت كلتا الطريقتين كفاءتهما للنموذج المستعمل في البحث، أخيرًا نأمل أن يشمل هذا النموذج تطبيقات أكثر استعمالاً في مجالات وطرائق تقدير مختلفة وخوارزميات ذكاء اخرى.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 2023/10/8
تاريخ قبول البحث: 2023/11/25
تاريخ رفع البحث على الموقع: 2024/6/30

الكلمات المفتاحية:
توزيع Gompertz المعمم، طريقة الخوارزمية الجينية، طريقة خوارزمية Nelder Mead، دالة المخاطرة، IMSE، محاكاة مونت كارلو .

اسم **للمراسلة:**
م.م. منال محمود رشيد
manal4254.@gmail.com

doi: https://doi.org/10.55562/jrucs.v55i1.629

1. المقدمة (Introduction) :

ان التطور ودخول الذكاء الاصطناعي الذي يشهده العالم خاصة في حقل العلم والتكنلوجيا ادى الى ظهور الكثير من الاجهزة الالكترونية والمعدات والمكائن المعقدة التي تستعمل في مجالات عديدة مثل الطب والهندسة وحقول الاتصالات والملاحة الفضائية وغيرها.

ترتبط قيمة النتائج العلمية المستخرجة بشكل مباشر بمدى دقة الاساليب والتجارب المستعملة في موضوع البحث وصحته ومن اجل صياغة الاساليب والتجارب في صورة تمكننا من الحصول على نتائج اكثر وضوحا وذات دقة عالية مقبولة ومرضية للتوصية منها، يفضل استعمال الاساليب الاحصائية المناسبة والاكثر دقة على وفق اسس علمية تكون حصيلتها اتخاذ القرار العلمي بقناعة علمية وبثقة كافية .

تلعب التوزيعات الاحصائية دوراً مهماً للغاية في الابحاث العلمية حيث انها مفيدة جداً في تحديد ووصف وتوقع ظواهر العالم الحقيقي، على الرغم من انشاء العديد من التوزيعات، الا ان هناك دائماً مجالاً لتطوير توزيعات اكثر مرونة لتناسب سيناريوهات العالم الحقيقي المحددة هنالك باحثون متحمسون يبحثون عن توزيعات جديدة واكثر مرونة نتيجة لذلك[1]. ومن اهم هذه التوزيعات هو توزيع Gompertz المعمم يعتبر واحدًا من أكثر النماذج شيوعًا وملائمة لبيانات النمو والبيانات الأخرى، وربما يأتي في المرتبة الثانية بعد النموذج اللوجستي (يُسمى أيضًا نموذج Verhulst) [7]. لقد قام الباحثون بتركيب نموذج Gompertz المعمم على كل شيء بدءًا من نمو النباتات، ونمو الطيور، ونمو الأسماك، ونمو الحيوانات الأخرى، وحتى نمو الورم ونمو البكتيريا [16-17]، والأدبيات العلمية الهائلة. يعد Gompertz المعمم حالة خاصة من نموذج ريتشاردز (Richards) ذو الاربع معلمات، تم اقتراح هذا النموذج، الذي كان يشار إليه في ذلك الوقت باسم قانون Gompertz النظري للوفيات، لأول مرة وتم تطبيقه قبل السيد Benjamin Gompertz في عام 1825 [18]. وقد قام بملاءمته مع العلاقة بين زيادة معدل الوفيات والعمر، وهو ما أشار إليه

"كمتوسط استنفاد قدرة الإنسان على تجنب الموت"، أو "الجزء المتبقي منه". القدرة على مقاومة التدمير وسرعان ما بدأت صناعة التأمين في استعمال أسلوبه في توقع مخاطر الوفاة. ومع ذلك، قدم Gompertz فقط دالة كثافة الاحتمال. Makeham [19] هو أول من ذكر هذا النموذج في شكله التراكمي المعروف، وبالتالي أصبح يعرف باسم نموذج Gompertz-Makeham [20] تم استعمال طريقة المربعات الصغرى لنموذج Gompertz للعثور على أفضل منحنى، على سبيل المثال[21] [22]. ومع ذلك، لم يقوموا بجعل النموذج خطيًا، كما حدث لاحقًا، ولكنهم فقط قاموا بتحويل القيم (المتغير التابع) لتسهيل تحديد مجموع المربعات. يبدو أن هذه الطريقة قد تم استعمالها حتى الأربعينيات [23]. عندما اقترح Hartley [24] وشرح لأول مرة كيفية جعل نموذج Gompertz خطيًا منذ العشرينيات من القرن العشرين، أصبح نموذج Gompertz-Makeham التراكمي هو المفضل في مجالات أخرى غير الوفيات البشرية، على سبيل المثال في التنبؤ بزيادة الطلب على السلع والخدمات، ومبيعات التبغ، ونمو حركة السكك الحديدية، والطلب على السيارات، [25]. كان Wright [26] أول من اقترح نموذج Gompertz للنمو البيولوجي، وربما كان Davidson [29] أول من طبقه على البيانات البيولوجية في دراسته لنمو كتلة الجسم في الماشية.

في عام 1931، درس كل من Weymouth وMcMillin وRich [30]عن نموذج Gompertz الذي نجح في وصف نمو حجم الصدفة في المحار، ودرس [Siliqua patula و Wey-mouth وThompson [31]عن الشيء نفسه بالنسبة لكوكل المحيط الهادئ، وسرعان ما بدأ الباحثون في ملاءمة النموذج لبياناتهم عن طريق الانحدار، وعلى مر السنين، أصبح نموذج Gompertz الشائع لنموذج الانحدار المفضل للعديد من أنواع نمو الكائنات الحية،

يتم استعمال العديد من إعادة المعلمات المختلفة لنموذج Gompertz التراكمي التقليدي. أحد هذه النماذج الأكثر أهمية هو ما اقترحه Zwietering وزملاؤه [32] لنمذجة النمو في عدد من البكتيريا وهو حاليًا أحد النماذج الأكثر شيوعًا في نمو الميكروبات [33]. هناك عملية إعادة صياغة بارزة أخرى لنموذج Gompertz وهي نموذج Gompertz-Laird ، الذي اقترحه Laird وتم تركيبه على بيانات نمو الورم [34]. يعتبر هذا النموذج مفيدًا بشكل خاص عندما نريد مناقشة القيمة الأولية (نقطة البداية على المحور السيني)، ويستخدم أيضًا لوصف النمو في الطيور والحيوانات، وخاصة الدواجن [35]. ومع ذلك، لا يمكن تفسير معلمات النموذج بسهولة دون تحويلها إلى قياسات أكثر فائدة. بالإضافة إلى إعادة معلمات Gompertz المتزايدة بشكل رتيبي .

2. هدف البحث (Purpose of Research)

إن الهدف الأساسي من هذا البحث هو الوصول إلى أفضل الطرائق في التقدير والحصول على نتائج أكثر وضوحا وذات دقة عالية مقبولة ومرضية. لذا تم مقارنة لبعض طرائق تقدير دالة المخاطرة لتوزيع (Gompertz المعمم) و المقارنة بتوظيف أسلوب المحاكاة باستخدام المقياس الإحصائي متوسط مربعات الخطأ التكاملي (IMSE) اخذين بنظر الاعتبار حجوم عينات مختلفة (صغيرة، متوسطة وكبيرة).

3. توزيع Gompertz المعمم (Generalized Gompertz Distribution)

يعد توزيع Gompertz المعمم واحدًا من أكثر النماذج شيوعًا وملائمة لبيانات النمو والبيانات الأخرى، وربما يأتي في المرتبة الثانية بعد النموذج اللوجستي (يُسمى أيضًا نموذج Verhulst) ويمكن إعطاء دالة probability density (PDF) function ودالة التوزيع التراكمية cumulative density function(CDF) لتوزيع GGD كالاتي [6]

$$f(x;\lambda,c,\theta)=\theta\lambda e^{cx}e^{-\frac{\lambda}{c(e^{cx}-1)}}\left[1-e^{-\frac{\lambda}{c(e^{cx}-1)}}\right]^{\theta-1} \quad (1)$$

$$F(x;\lambda,c,\theta)=\left[1-e^{-\frac{\lambda}{c(e^{cx}-1)}}\right]^{\theta} \quad (2)$$

اذ ان: $x>0,\lambda,\theta>0,c\geq 0$:

4. دالة المخاطرة (Hazard function) :

يعرف بأنه احتمال فشل المفردة خلال الفترة $(x, x+\Delta x)$، علماً أن المفردة لم تفشل حتى الوقت x. ويرمز لدالة معدل الفشل بالرمز $h(x)$، وان:

$$h(x)=\lim_{\Delta x\to 0}\frac{Pr[x<X\leq x+\Delta x\backslash X>x]}{\Delta x} \quad (3)$$

$$h(x)=\frac{1}{S(x)}\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x} \quad (4)$$

$$h(x)=\frac{1}{S(x)}\cdot\frac{dF(x)}{dx} \quad (5)$$

$$h(x)=\frac{f(x)}{S(x)} \tag{6}$$

$$h(x)=\frac{\theta\lambda e^{cx}e^{-\frac{\lambda}{c(e^{cx}-1)}}\left[1-e^{-\frac{\lambda}{c(e^{cx}-1)}}\right]^{\theta-1}}{1-\left[1-e^{-\frac{\lambda}{c(e^{cx}-1)}}\right]^{\theta}} \tag{7}$$

5. طرائق التقدير (Estimation Methods):

5.1. طريقة الامكان الاعظم (Method of Maximum likelihood Probability) :

إذا كانت $X=(X_1,X_2,\dots\ X_r)$ عبارة عن عينة عشوائية مستقلة تمتلك دالة PDF من التوزيع GGD، فإن الدالة الاحتمالية والدالة اللوغارتمية يتم كتابتها بالشكل الاتي وعلى التوالي:

$$L(x)=\prod_{i=1}^{r} f(x_i)=\theta^r\lambda^r e^{\sum_{i=1}^{r}cx_r}e^{-\frac{\lambda}{c(e^{\sum_{i=1}^{r}cx_r}-r)}}\left[r-e^{-\frac{\lambda}{c(e^{\sum_{i=1}^{r}cx_r}-r)}}\right]^{\theta-r} \tag{8}$$

$$\ln L(x)=r\ln\theta+r\ln\lambda+\sum_{i=1}^{r}cx_r-\lambda+c(e^{Cx_r}-r)+(\theta-r)ln\left[r-e^{-\frac{\lambda}{c(e^{\sum_{i=1}^{r}cx_r}-r)}}\right] \tag{9}$$

$$\frac{dLnL}{d\hat{\lambda}}=\frac{r}{\hat{\lambda}}-\frac{1}{\hat{c}}\sum_{i=1}^{r}e^{\hat{C}x_r-r}[1+r-\hat{\theta}]=0 \tag{10}$$

$$\frac{dLnL}{d\hat{c}}\sum_{i=1}^{r}x_r-\frac{\hat{\lambda}}{\hat{c}}\sum_{i=1}^{r}e^{\hat{C}x_r-r}\left[\sum_{i=1}^{r}x_r-\frac{1}{\hat{c}}+(\hat{\theta}-r)\sum_{i=1}^{r}x_r+\frac{\hat{\theta}-r}{\hat{c}}\right]=0 \tag{11}$$

$$\frac{dLnL}{d\hat{\theta}}=\frac{r}{\hat{\theta}}+\ln r+\frac{\hat{\lambda}}{\hat{c}}\sum_{i=1}^{r}e^{\hat{C}x_r-r}=0 \tag{12}$$

من خلال المعادلات (12,11,10) فيما يتعلق بالمعلمات الثلاث ومحاولة حلها للحصول على المعلمات المقدرة في شكل صريح أمر معقد، لذلك سيتم الحصول على هذه التقديرات بالطرق العددية ومن ثم تعويضها في دالة المخاطرة لنحصل على مقدرات هذه الدالة معادلة رقم (7).

$$\hat{h}(x)=\frac{\hat{\theta}\hat{\lambda}e^{\hat{c}x}e^{-\frac{\hat{\lambda}}{\hat{c}(e^{\hat{c}x}-1)}}\left[1-e^{-\frac{\hat{\lambda}}{\hat{c}(e^{\hat{c}x}-1)}}\right]^{\hat{\theta}-1}}{1-\left[1-e^{-\frac{\hat{\lambda}}{\hat{c}(e^{\hat{c}x}-1)}}\right]^{\hat{\theta}}} \tag{13}$$

5.2. طريقة الخوارزمية الجينية (Genetic algorithm method)

هي إحدى طرائق الذكاء الاصطناعي والتي يمكن تعريفها للحصول على أفضل حل لمسألة قيد الدراسة بطريقة فعالة وسريعة. جاءت فكرتها من أستاذ في علوم الحاسوب جون هولاند (John Holland)، والتي تهدف إلى تحديث مفهوم عملية التطور الطبيعي وتصميم أنظمة صناعية ذات خصائص مماثلة للأنظمة الطبيعية وطموحه المستمر لتحسين أداء الأنظمة الحسابية جعل الخوارزميات الجينية أكثر فعالية. في حل مشكلات التحسين[1-3].

- **تطبيق مراحل الخوارزمية الجينية لتوزيع (GGD) :**

نطبق مراحل الخوارزمية الجينية في معادلة دالة الهدف لكل طريقة لإيجاد تقديرات معلمات لتوزيع (GGD)وفقاً لما يأتي [27-28]:

1. البداية: تكوين الكروموسوم من خلال قيم β_p التي تشكل جينات الكروموسوم وان (P=0,1,2,...,p) ضمن الاعداد الحقيقية .
2. التهيئة: انشاء الجيل الابتدائي عن طريق ايجاد قيمة اولية للجينات مع القيم العشوائية لمجموعة القيود الاخرى .
3. دالة الهدف: يتم تقييم الكروموسوم من حيث الكفاءة وصولاً الى الحل الامثل بتحديد قيمة β_p
4. اجراء عملية الاختبار للكرموسوم الذي يمتلك قيمة دالة الهدف صغيرة باختيار الاحتمال الكبير لها وايجاد دالة التقييم له من خلال المعادلة التالية

$$Fitness\ Function = \frac{1}{1 + objective\ function} \tag{14}$$

من خلال صيغة دالة التقييم نستطيع ايجاد احتمالية هذه الدالة (افضل القيم) حسب الصيغة الرياضية الاتية[37] :

$$C_{(i)} = \frac{f(i)}{\sum_{i=1}^{n} f(i)} \tag{15}$$

اذ ان :

$C_{(i)}$: تمثل احتمالية الفرد i

$f(i)$: دالة التقييم الفرد i

n : يمثل حجم المشاهدات

1. في هذه الخطوة يتم تهجين للكروموسومات الجيدة في صفاتها عن طريق التزاوج بين كل كروموسومين وينطبق احد معاييره وهو التهجين المنظم بالاعتماد على احتمالية التهجين Pc وتقارن هذه القيمة مع قيمة الجينات للكرموسومين (الاباء) لتكوين الجيل الجديد الابناء ويحدث التبادل عندما تكون قيمة الجين الاكبر او تساوي القيمة الاحتمالية [38].
2. اخر خطوة يمكن ان تمر بها الكروموسومات هي عملية الطفرة وايضا تعتمد على مقدار احتمالي (Pm) للمعلمات استبدال جينات منتقاة عشوائياً مع قيمة جديدة ايضاً حصلنا عليها بشكل عشوائي وفق الصيغة التالية [36]:

مجموع الجينات = عدد الجينات في الكروموسوم * مجموع السكان

5.3. طريقة الخوارزمية Nelder Mead (Nelder Mead algorithm method)

هي خوارزمية بحث عددية تستعمل العلاقات الهندسية لإيجاد افضل تقريب للحد الادنى لدالة الهدف في فضاء متعدد الابعاد، تم صياغتها واستعمالها لأول مرة من قبل الباحثين (Roger Mead & john Nelder) في عام 1965 ولقد تم تسميتها بطريقة (Downhil) ومن اهم مميزاتها في خوارزمية (NM) يتم استعمال الدالة فقط دون الحاجة الى مشتقات الدالة وتعتبر ايضاً اكثر كفاءة من الطرق التقليدية البديلة ويتم استعمالها في العديد من المشاكل الرياضية المشابهة عند مواجهة معادلات غير خطية او دوال معقدة عند اشتقاق الدالة وتعتمد الخوارزمية على مفهوم (Simplex) وهو عبارة عن قيم اختيارية او عشوائية لمعلمات التوزيع او الدالة قيد الدراسة وهو شكل هندسي (n) من الابعاد لأحجام الغير صفرية (n+1) من النقاط او القمم (الرؤوس). [4-39]

تتم الخوارزمية عن طريق اعداد (Simplex) الاولي وان طريقة (NM) تعتمد بشكل كبير على الافتراض الاولي (التقديرات الاولية لمعلمات التوزيع) لإنتاج (Simplex الاولي) [40]

الية عمل الخوارزمية عن طريق ترتيب النقاط ترتيباً تصاعدياً

$$T_1, T_2, T_3, \dots, T_{n+1} \qquad ,T = (\lambda, c, \theta) \tag{16}$$

وتعويض كل نقطة بدالة الهدف الموضوعة

$$h(T_1) < h(T_2) < \dots < h(T_{n+1}) \tag{17}$$

وتشير T_1 الى افضل نقطة (القيم الافضل للمعلمة T) و $h(T_1)$ الى القيمة الافضل لدالة الهدف , بينما T_{n+1} تعني قيمة ادنى نقطة و $h(T_{n+1})$ القيمة الادنى لدالة الهدف (القيم الادنى للمعلمة T) [40]

تمتد هذه القيم الى سلسلة مستمرة بأسلوب تكراري خلال مجموعة من العمليات لانعكاس (Reflection ,r) والتوسع (Expansion, e) والانكماش(Contraction, c) والتقلص (Shrink, Sh) وتتم في العمليات السابقة استبدال أسوأ نقطة بنقطة جديدة افضل منها ويستمر التكرار وصولاً الى الحد الادنى الامثل لدالة الهدف الذي يعتبر الحل الامثل . [41]

- **طريقة خوارزمية Nelder Mead لتقدير معلمات أنموذج GGD :**

اقترح الباحث في هذا البحث توظيف Nelder Mead (NM) لتقدير معلمات انموذج GGD (λ, c, θ) وفيما يلي خطوات تقدير المعلمات وهي الخطوات العامة لـ (NM) مع التعديل بما يتناسب مع تقدير معلمات قيد الدراس[41-42]:

1. تحديد دالة الهدف h(T) التي تتضمن المعلمات (λ, c, θ) ونهدف في هذه الخوارزمية لجعل الدالة اقل ما يمكن وهنا تكون دالة الهدف وهو سالب لوغارتم دالة الامكان :

$$h(T) = -Ln(T), \text{ where } T = ((\lambda, c, \theta) \qquad (18)$$

2. تحديد قيم معلمات الخوارزمية ($1\zeta=.2$, $\varepsilon=3$, $\delta=0.5$, $\gamma=1.3$) المرافقة لعمليات (الانعكاس، التوسع، الانكماش، التقلص) وكذلك تحديد عدد الحلول (z+1) [5]
3. توليد مصفوفة الحل الاولي، المصفوفة V بأبعاد (z+1)*3)) حيث تمثل الصفوف i عدد الحلول الممكنة لكل معلمة وتمثل الاعمدة j المعلمات المراد تقديرها

$$V=\begin{bmatrix} \lambda_1 & \theta_1 & c_1 \\ \vdots & \vdots & \vdots \\ \lambda_{z+1} & \theta_{z+1} & c_{z+1} \end{bmatrix}$$

يمكن التعبير عن كل صف في المصفوفة V بالرمز T والذي يمثل المعلمات المدروسة فتصبح المصفوفة كالاتي :

$$V=\begin{bmatrix} T_1 \\ \vdots \\ T_{z+1} \end{bmatrix} \quad , where\ T = (\lambda, c, \theta)$$

4. احتساب دالة الهدف لكل صف في (V) :

$$h(T_1), h(T_2), \dots, h(T_{z+1}) \qquad (19)$$

5. يتم ترتيب الحلول (القمم) في المصفوفة V حسب قيمة دالة الهدف من اقل قيمة الى اعلى قيمة وتكون اعلى قيمة من دالة الهدف هي التي تمتلك اسوأ حل :

$$h(T_1) < h(T_2) < \dots < h(T_{z+1}) \qquad (20)$$

6. احتساب متوسط المصفوفة (v) كالاتي :

$$\bar{V} = \frac{\sum_{i=1}^{z} T_i}{z} \quad , where\ i = 1,2,3,\dots,z \qquad (21)$$

7. احتساب نقطة الانعكاس (r) التي هي نقطة اختبار جديدة ويرمز لها T_r وتحسب وفق الصيغة الاتية :

$$T_r = \bar{V} + \zeta(\bar{V} - T_{z+1}) \qquad (22)$$

ثم احتساب دالة الهدف ($h(T_r)$) ضمن الافتراضات الاتية :

- اذا كانت $h(T_1) < h(T_r) < h(T_z)$ يتم جعل $(T_r = T_{z+1})$ والذهاب الى الخطوة (11)
- اذا كانت $h(T_r) < h(T_1)$ يتم الانتقال الى الخطوة (8).

8. ايجاد نقطة اختبار جديدة وهي التوسع (e) ويرمز لها (T_e) وفق الصيغة الاتية :

$$T_e = \bar{V} + \varepsilon(T_r - \bar{V}) \qquad (23)$$

ثم احتساب دالة الهدف $(h(T_e))$ ضمن الافتراضات الاتية :

- اذا كانت $h(T_e) < h(T_r)$ يتم جعل $(T_e = T_{z+1})$ والذهاب الى الخطوة (11)
- عد ذلك اذا كانت $h(T_e) \geq h(T_r)$ يتم الانتقال الى الخطوة (9)

9. ايجاد نقطة اختبار جديدة وهي الانكماش (c) ويرمز لها (T_c) وفق الصيغة الاتية :

$$T_c = \bar{V} + \delta(T_{z+1} - \bar{V}) \qquad (24)$$

يتم احتساب دالة الهدف $(h(T_c))$ ضمن الافتراضات الاتية :

- اذا كانت $h(T_c) < h(T_z)$ يتم جعل $(T_c = T_{z+1})$ والذهاب الى الخطوة (11)
- عند ذلك يتم الانتقال الى الخطوة (10)

10. ايجاد نقطة اختبار جديدة وهي التقلص (sh) ويرمز لها (T_{sh}) اذ يتم التقلص نحو افضل مرشح للحل وتحسب وفق الصيغة الاتية :

$$T_{sh} = T_1 + \gamma(T_i - T_1) \tag{25}$$

11. اذا تحقق شرط التوقف الذي يحسب من خلال الصيغة الاتية :

$$\left|\frac{\max(h)-\min(h)}{\max(h)}\right| < \varepsilon \text{ -1} \tag{26}$$

اذ ان (ε) عددا صغيرا جداً

يتم طباعة الحل الامثل , عدا ذلك يتم الرجوع الى الخطوة (7)

12. النهاية

6. الجانب التجريبي

تم استعمال اسلوب المحاكاة بطريقة مونت كارلو للمقارنة بين طرائق التقدير المختلفة حيث يتميز هذا الاسلوب بالمرونة ويوفر الكثير من التكاليف عن طريق اخذ بنظر الاعتبار حجوم العينات المختلفة والقيم المختلفة لمعلمات التوزيع وتكرار التجربة في كل مرة ويتم في هذا الاسلوب توليد البيانات دون اللجوء الى البيانات الحقيقية مع عدم الاخلال بالدقة المطلوبة وتتلخص هذه الطريقة بالخطوات التالية[2]

1. تحديد القيم الافتراضية :حيث تم اختيار خمسة حجوم عينات وهي (10,20,35,80,150) وكذلك تم استعمال قيم افتراضية لثلاث معلمات.
2. توليد البيانات:

- اذ تم توليد المتغير العشوائي بطريقة التحويل المعكوس للدالة كما يلي:

$$x = \frac{1}{c}\ln\left(\frac{\lambda \ln y^{\frac{1}{\theta}}}{c} + 1\right) \tag{27}$$

3. ترتيب البيانات التي تم توليدها ترتيباً تصاعدياً $X_1 < X_2 < \cdots < X_i$
4. حل المعادلات التي تم الوصول اليها بالطرق العددية
5. تم تحديد الطريقة الافضل عن طريق مقياس المقارنة متوسط مربعات الخطأ التكاملي (IMSE) في حالة تقدير دالة المخاطرة

$$IMSE(\hat{h}(x)) = \frac{1}{r}\sum_{i=1}^{r}\left[\frac{1}{n_t}\sum_{j=1}^{n_t}\left(\hat{h}_i(x_j) - h(x_j)\right)^2\right] \tag{28}$$

اذ أن :

r : عدد تكرارات التجربة (1000) مرة

n_t : عدد البيانات المولدة لكل عينة

$h(x_j)$: دالة المخاطرة حسب القيم الابتدائية وعلى التوالي .

وقد تم الحصول على النتائج المتمثلة بالجداول الاتية :

جدول (1): يبين قيم معالم الافتراضية والمقدرة لـ (λ, c, θ) وحجوم العينات ولجميع التجارب

		$\hat{\lambda}$		$\hat{c}$		$\hat{\theta}$	
model		G.E(ml)	(ml)NM	G.E(ml)	(ml)NM	G.E(ml)	NM(ml)
λ =0.2,C=0.4,θ=2.1		0.100015	0.999995	2.490443	2.681577	0.200009	0.999987
		0.100014	1.000568	2.399779	2.794981	0.199988	1.000013
		0.099976	0.999996	2.484991	2.788445	0.200011	0.999991
		0.100005	0.99999	2.709721	2.780355	0.199989	1.000007
		0.100092	1.000002	2.682198	2.093172	0.199997	0.999986

λ =1.4,C=1.3,θ=1.0		0.300011	0.999995	2.040121	3.000748	0.100003	1.50001
		0.300026	1.000009	32.02206	3.000722	0.100013	1.500013
		0.299983	1	3.49318	3.499147	0.100002	1.500006
		0.300014	1.000007	2.495434	3.499768	1.000001	1.500001
		0.300014	0.999994	2.525632	3.497882	0.100015	1.499999
λ =4.3,C=4.1,θ=3.4		0.200008	1.500036	2.05412	2.000036	0.300007	1.500002
		0.200005	1.499999	1.091727	2.0018	0.299992	1.499991
		0.199991	1.500006	1.088246	1.999153	0.299991	1.499995
		0.199997	1.499982	4.003771	4.000637	0.300006	1.500018
		0.199987	1.500011	3.060238	3.999708	0.29997	1.50002
λ =6.7,C=5.1,θ=5.6		1.999996	0.100017	2.997847	3.9997705	2.000007	0.200001
		1.999995	1.00E-01	2.000657	4.4987679	2.000003	0.199994
		2.000009	0.099986	2.001963	4.4981918	2.000008	0.200004
		2	0.100018	4.999174	4.4993785	2.000011	0.2
		2.000022	1.00E-01	4.001468	4.0011797	1.999987	0.199993

جدول (2): يبين قيم IMSE لتقدير دالة المخاطرة بجميع الطرائق وحجوم العينات ولجميع التجارب

Model	N	G.E(ML)	(ML)NM	Best
λ =0.2,C=0.4,θ=2.1	10	4.71E-08	2.87E-11	(ML)NM
	20	8.33E-09	5.78E-11	(ML)NM
	35	7.10E-09	2.11E-10	(ML)NM
	80	2.81E-09	2.40E-10	(ML)NM
	150	6.99E-10	1.28E-12	(ML)NM
λ =1.4,C=1.3,θ=1.0	10	1.23E-08	1.81E-11	(ML)NM
	20	6.58E-09	7.32E-11	(ML)NM
	35	2.84E-09	1.01E-10	(ML)NM
	80	2.87E-09	2.80E-10	(ML)NM
	150	5.61E-10	1.20E-15	(ML)G. E
λ =4.3,C=4.1,θ=3.4	10	2.75E-11	1.07E-08	G. E(ML)
	20	6.48E-11	4.17E-09	G. E(ML)
	35	1.07E-10	1.92E-09	G. E(ML)
	80	1.79E-10	1.64E-09	G. E(ML)
	150	3.19E-10	1.35E-09	G. E(ML)
λ =6.7,C=5.1,θ=5.6	10	5.50E-12	1.81E-08	G. E(ML)
	20	1.45E-11	3.86E-09	G. E(ML)
	35	5.88E-11	3.38E-09	G. E(ML)
	80	8.79E-11	1.95E-09	G. E(ML)
	150	8.73E-11	9.01E-10	G. E(ML)

يوضح جدول رقم (2) قيم متوسطات مربعات الخطأ التكاملي (IMSE) لمقدرات دالة المخاطرة لطريقتي الخوارزمية الجينية وخوارزمية Nelder Mead وكالاتي :

1. نلاحظ بالنموذج الاول عندما كانت (λ =0.2,C=0.4,θ=2.1) وعند حجوم العينة n=10,20,35,80,150 ظهرت طريقة خوارزمية Nelder Mead بالأفضلية
2. ظهر في النموذج الثاني عندما كانت (λ =1.4,C=1.3,θ=1.0) وعند حجوم العينة n=10,20,35,80 اثبتت خوارزمية Nelder Mead كفاءتها اما عند حجم عينة n=150 اثبتت الخوارزمية الجينية بالأفضلية

3. ظهر في النموذج الثالث والرابع عندما كانت (λ =4.3,C=4.1,θ=3.4) و(λ =6.7,C=5.1,θ=5.6)ولجميع حجوم العينات اثبتت الخوارزمية الجينية بالأفضلية.

7. الاستنتاجات (Conclusions)

من خلال نتائج جانب المحاكاة توصلنا الى الاستنتاجات الاتية :

1. تم ايجاد مقدرات معالم (λ,c,θ) لأنموذج Gompertz المعمم لحجوم العينات n=10,20,35,80,150 لطريقتي الخوارزمية الجينية وخوارزمية Nelder Mead
2. وجد بالنموذج الاول عندما كانت (λ =0.2,C=0.4,θ=2.1) ولجميع حجوم العينات ظهرت طريقة خوارزمية Nelder Mead بالأفضلية لدالة المخاطرة باستعمال متوسط مربعات الخطأ التكاملي (IMSE)
3. وجد في النموذج الثاني عندما كانت (λ =1.4,C=1.3,θ=1.0) وعند حجم عينة n=10,20,35,80 اثبتت خوارزمية Nelder Mead كفاءتها اما عند حجم عينة n=150 اثبتت الخوارزمية الجينية بالأفضلية
4. وجد في النموذج الثالث والرابع عندما كانت (λ =4.3,C=4.1,θ=3.4) و(λ =6.7,C=5.1,θ=5.6)ولجميع حجوم العينات اثبتت الخوارزمية الجينية بالأفضلية .

8. التوصيات و الدراسات المستقبلية (Recommendations and future studies)

من خلال ما تم التوصل اليه من الاستنتاجات نوصي بالآتي :

1. اعتماد الخوارزمية الجينية وخوارزمية Nelder Mead في ايجاد مقدرات دالة البقاء ودالة الكثافة الاحتمالية لأنموذج (Gompertz المعمم).
2. ضرورة اعتماد خوارزميات اخرى لذكاء الاصطناعي غير التي ذكرت في البحث في حالة حجوم العينات الصغيرة المتوسطة والكبيرة.
3. بإجراء بحوث مستقبلية باستعمال طريقة بيز وطريقة التجزيئية لتقدير دالة المخاطرة لأنموذج (Gompertz المعمم) لثلاث معالم في حالة البيانات المفقودة وكذلك في حالة البيانات تحت المراقبة وبشكل مفصل .
4. استعمال مبدأ مجموعة المعاينة المرتبة (Rank Set Sampling) لتقدير دالة المخاطرة لتوزيع (Gompertz المعمم) ومقارنته مع طرائق التقدير الاعتيادية .

References

[1] Rashid, Manal Mahmoud (2021) “Employing the artificial intelligence algorithm in Comer’s generalized beta estimators and comparing them to classical methods with a practical application” M. Sc. thesis in statistics, College of Administration and Economics, University of Baghdad.

[2] Al-Amiri, Samah Sabah (2021) “Estimating survival and risk functions using log-logistic distribution statistics with a practical application” M. Sc. thesis in statistics, College of Administration and Economics, University of Baghdad.

[3] Al-Rudaini, Sarah Adel Mazloum (2019) “Using the genetic algorithm to estimate the parameters of the binary logistic regression model with a practical application” M. Sc. thesis in statistics, College of Administration and Economics, University of Baghdad.

[4] Al-Khafaji, Murtada Alaa Abdel-Yama (2021) “Comparison of some estimators of the twisted normal distribution parameters using genetic algorithms” M. Sc. thesis in statistics, College of Administration and Economics, University of Baghdad.

[5] Fadaam, Intisar, Muhammad, Nour (2017), “Comparison of some methods for estimating the parameters of the complex Poisson-Weibull distribution,” Iraqi Journal of Statistical Sciences, Volume (23), Issue (101)

[6] Mohammed Obeidat , Amjad Al-Nasser& Amer I. Al-Omari , (2020),"Estimation of Generalized Gompertz Distribution Parameters under Ranked-Set Sampling", Journal of Probability and Statistics ,Volume 2020, Article ID 7362657.
https://doi.org/10.1155/2020/7362657

[7] Kathleen M. C. Tjørve, Even Tjørve, (2017), "The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family", PLoS ONE 12(6): e0178691.
https://doi.org/10.1371/journal. pone.0178691

[8] B. Gompertz, “On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies,” Philosophical Transactions of the Royal Society of London, Vol. 115, pp. 513–583, 1825.

[9] H. M. Moustafa and S. G. Ramadan, “Errors of misclassification and their probability distributions when the parent populations are gompertz,” Applied Mathematics and Computation, Vol. 163, No. 1, pp. 423–442, 2005.

[10] M. M. M. El-Din, Y. Abdel-Aty, and M. H. Abu-Moussa, “Statistical inference for the gompertz distribution based on type-ii progressively hybrid censored data,” Communications in Statistics - Simulation and Computation, Vol. 46, No. 8, pp. 6242–6260, 2017.

[11] A. El-Gohary, A. Alshamrani, and A. N. Al-Otaibi, “The generalized gompertz distribution,” Applied Mathematical Modeling, Vol. 37, No. 1-2, pp. 13–24, 2013.

[12] P. Borges, “EM algorithm-based likelihood estimation for a generalized gompertz regression model in presence of survival data with long-term survivors: an application to uterine cervical cancer data,” Journal of Statistical Computation and Simulation, Vol. 87, No. 9, pp. 1712–1722, 2017.

[13] E. Demir and B. Saracoglu, “Maximum likelihood estimation for the parameters of the generalized gompertz distribution under progressive type-ii right censored samples,” Journal of Selcuk University Natural and Applied Science, Vol. 4, No. 1, pp. 41–48, 2015.

[14] E. A. Ahmed, “Estimation of some lifetime parameters of generalized gompertz distribution under progressively type-ii censored data,” Applied Mathematical Modeling, Vol. 39, No. 18, pp. 5567–5578, 2015.

[15] H. H. Abu-Zinadah and A. S. Al-Oufi, “Different method estimations of the three parameters of exponentiated gompertz distribution,” Applied Mathematics & Information Sciences, Vol. 10, No. 2, pp. 705–710, 2016.

[16] Winsor CP." The Gompertz curve as a growth curve". Proc. Nat. Acad. Sci. 1932; 18(1):1-8. PMID: 16577417

[17] Halmi MIE, Shukor MS, Johari WIW, Shuker MY. “Evaluation of several mathematical models for fitting the growth of the algae Dunaliella tertiolecta”. Asian Journal of Plant Biology. 2014; 2(1):1-6.

[18] Gompertz B. "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies", Philosophical Transactions of the Royal Society of London B: Biological Sciences. 1825; 182:513-85.

[19] Makeham WM. “On the integral of Gompertz's function for expressing the values of sums depending upon the contingency of life”. Journal of the Institute of Actuaries and Assurance Magazine. 1873; 17(5):305-27.

[20] Greenwood M. “Discussion on the value of life-tables in statistical research”. Journal of the Royal Statistical Society. 1922; 85(4):537-60.

[21] Rietz HL. “Properties of Makeham's laws of mortality with applications”. The American Mathematical Monthly. 1921; 28(4):158-65.

[22] Trachtenberg HL. “The wider application of the Gompertz law of mortality”. Journal of the Royal Statistical Society. 1924; 87(2):278-90.

[23] Stoner PM. “Fitting the exponential function and the Gompertz function by the method of least squares”. Journal of the American Statistical Association. 1941; 35(216):515-8.

[24] Hartley HO. “The estimation of non-linear parameter by internal least squares”. Biometrica. 1948; 35(1/2):32-45.

[25] Proscott RB. “Law of growth in forecasting demand”. Journal of American Statistical Association. 1922; 18(140):471-9.

[26] Wright S. Book review. American Statistical Society Quarterly Journal. 1926; 21:493-7.

[27] Sefian , S., Benbouziane , M. ,(2012) ,"Portfolio Selection Using Genetic Algorithm", MPRA Paper No. 41783, pp. 1-12

[28] Mahdavi, I., Paydar, M., & et al., (2008),"Genetic algorithm approach for solving a cell formation problem in cellular manufacturing", Expert Systems with Applications, Volume 36, Issue 3, Part 2

[29] Avidson FA. Growth and senescence in purebred Jersey cows. Univ. of Illinois Agricultural Experiment Station Bull. 1928; 302:192-9

[30] Weymoth FW, McMillin HC, Rich WH. “Latitude and relative growth in razor clam Siliqua patula”. J Exp. Biol. 1931; 8:228-49.

[31] Weymouth FW, Thompson H. S. The age and growth of the pacific cockle (Cardium corbis, Martyn). Bulletin US Bureau of Fisheries. 1931; 46:633-41.

[32] Zwietering MH, Jongenburger I, Rombouts FM, Van T Rie T K. “Modeling of the bacterial growth Curve”. Appl. Env. Micriobiol. 1990; 56(6):1975-81

[33] Skinner GE, Larkin JW. “Mathematical modeling of microbial growth: a review”. Journal of Food Safety. 1994; 14:1975-217.

[34] Laird AK. “Dynamics of tumor growth”. British Journal of Cancer. 1964; 18:490-502.

[35] Marinho KMdS, de Freitas AR, Falcão J, de S., Dias FEF. “Nonlinear models for fitting growth curves of Nellore cows reared in the Amazon biome”. Revista Brasileira de Zootecnia. 2013; 42(9):645±50

[36] Goldberg, D. E., & Deb, K., (1991), "A Comparative Analysis of Selection Schemes Used in Genetic Algorithms, Foundation of Genetic Algorithms", San Francisco: CA: Morgan Kaufmann, pp. 69-93.

[37] Hadji, S. & et al., (2015), "Theoretical and experimental analysis of genetic algorithms based MPPT for PV systems", Energy Procedia, 74, pp. 772-787.

[38] Liu, H., Ong, C., (2008), "Variable selection in clustering for marketing segmentation using genetic algorithms", Expert Systems with Applications, 34(1):502-510

[39] Barati , Reza (2011), “Parameter Estimation of Nonlinear Muskingum Models Using Nelder-Mead Simplex Algorithm”, Journal Of Hydrologic Engineering, Vol. 15, No. 11.

[40] Lasheen, Atef , El-Garhy, Ahmed , And Saad , Elsayed (2009), “Using Hybrid Genetic and Nelder-Mead Algorithm for Decoupling of MIMO Systems with Application on Two Coupled Distillation Columns Process”, International Journal Of Mathematics and Computers in Simulation, Vol 3, No. 3

[41] Kiliç , Muhammet (2020) , "Using Genetic Algorithms For Parameter Estimation Of A Two-Component Circular Mixture Model", 4th International Conference on Computational Mathematics and Engineering Sciences (CMES-2019)

[42] Yalçinkaya , Abdullah , Şenogˇlu , Birdal And Yolcu , Ufuk (2018),"Maximum Likelihood Estimation For The Parameters Of Skew Normal Distribution Using Genetic Algorithm", Swarm And Evolutionary Computation, Vol. 38, pp. 127-138.

Journal of AL-Rafidain University College for Sciences (2024); Issue 55; pp. 33-43

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Comparison of Some Methods to Estimate the Hazard Function of the Generalized Gompertz Model Using Simulation

Lect. Assist. Manal M. Rashid

manal4254.@gmail.com

General Directorate of Education Baghdad - Al-Rusafa , Baghdad, Iraq

Article Information

Article History:
Received: October, 8, 2023
Accepted: November, 25, 2023
Available Online: June, 30, 2024

Keywords:
Generalized Gompertz distribution, Genetic algorithm method, Nelder Mead algorithm method, Hazard function, IMSE, Monte Carlo simulation.

Correspondence:
Manal M. Rashid
manal4254.@gmail.com

Abstract

This study examined the generalized Gompertz distribution, one of the most significant and popular failure models in reliability and life tests when the population is non-homogeneous. Two significant techniques for estimating the hazard function—the genetic method and the Nelder Mead algorithm—were employed once this model's significance, applications, and construction were thoroughly covered. The Monte Carlo method was utilized to simulate and compare these approaches because of its flexibility and cost-effectiveness in accounting for varying sample sizes (large, medium, and small) as well as predetermined values for the distribution and frequency parameters. Every time, try something new. This approach generates the data without using actual data. For the research model, both approaches have been shown to be adequate. Lastly, we anticipate that this model will incorporate additional intelligence algorithms, estimation techniques, and applications that are more broadly utilized across other domains.

doi: https://doi.org/10.55562/jrucs.v55i1.629