# إستخدام الخوارزمية الجينية لتقدير معلمة الحامض النووي الرايبي (RNA)

# Using Genetic Algorithm for Estimating the Parameter of (RNA)

* د. بان احمد حسن متراس ** فرح سعد نشاط

## الملخص

تم في هذا البحث استخدام الخوارزمية الجينية لتقدير المعلمـة θ الموجـودة فـي توزيـع بولتزمان الذي يخضع له تركيب الحامض النووي الرايبي (RNA) إذ تم اقتراح خوارزميتان أولهما تقوم بإيجاد قيمة المقدر الذي يعظم دالة الترجيح لتوزيع بولتزمان، والثانية تقوم بتقليل دالة قيد الجيل لتوزيع بولتزمان باستخدام الخوارزمية الجينية، ولقد استخدم نظـام Matlab (7.0) في كتابة برامج الخوارزميات والتي حصلنا من خلالها على النتائج التالية: أعظم قيمة لمقدر الترجيح لتوزيع بولتزمان ظهرت عند القيمة 94.1614- حيث كانـت قيمـة θ هـي 0.1457، وأقـل قيمـة لدالـة قيـد الجيـل لتوزيـع بولتزمـان ظهـرت عنـد القيمـة $0.951039101*10^7$ حيث كانت قيمة θ هي 4.4066-

## Abstract

In this paper the genetic algorithm has been used to estimate the parameter θ which exist in Boltzmann Distribution which controls the structure of the Ribo Nucleic Acid (RNA). Two algorithms have been suggested. The first found the value of the estimator which maximizes the likelihood function of Boltzmann Distribution. The second minimized the generation constraint of Boltzmann Distribution by using the genetic algorithm. Matlab (7.0) has been used in writing the programs of algorithms and achieved the following results: The maximum value for the likelihood estimator for Boltzmann Distribution appear at the value -4.1614 where the value of θ is 0.1457, and the minimum value for the Constraint Generation for Boltzmann Distribution appear at the value $0.951039101*1^7$ where the value of θ is -4.4066.

* استاذ مساعد/قسم بحوث العمليات والتقنيات الذكائية/كلية علوم الحاسوب والرياضيات.
** مدرس مساعد/جامعة دهوك.

## 1. مقدمة Introduction

شهد القرن العشرين تقدماً هائلا في الأساليب العلمية المستخدمة في البحث العلمي في حقول المعرفة كافة ويشمل هذا التطور علم الإحصاء الذي أصبح علماً مـستقلا لـه أهميتـه بوصفه وسيلة للبحث العلمي وأداته، ونظراً لأهميته واستخدامه في العديد من مجالات الحياة، فقد اهتم العديد من الباحثين به وقاموا بدراسته وادخاله في بعض المجالات الحياتيـة ومنهـا دراسة بعض الأحماض النووية التي تدخل في التركيب الداخلي لجسم الإنسان ومن أهمها الـ RNA والـ DNA، اذ أن سلاسل هذه الأحماض تعتبر متغيرات عشوائية تخضع لتوزيعات احتمالية [7] [6].

والمقصود بتوزيع المتغير العشوائي (أو التوزيع الإحتمالي لمتغيـر عـشوائي) بأنـه تعريف السلوك الرياضي لإحتمالات قيم المتغير. أي أن التوزيع الإحتمالي للمتغير العـشوائي هو القانون الإحتمالي لذلك المتغير [2].

أن لنظرية التقدير أهمية كبيرة في تطبيقات النظرية الإحصائية في الجوانب العلميـة، إذ أنها توفر قواعد يتم بموجبها تقدير معلمات مجهولة لظاهرة معينـة فـي الواقـع العملـي ولمختلف المجالات.

ويمكن تجزئتها إلى جزئين متكاملين، الأول يهتم بالبحث عن أفـضل تقـدير لمعلمـة مجهولة في المجتمع ويسمى بالتقدير النقطي (Point Estimation)، ويهتم الجـزء الثـاني بالبحث عن أفضل فترة يمكن حصر قيمة المعلمة المجهولة خلالها وهذا ما يسمى بتقدير الفترة (Interval Estimation) [5].

لقد كان هناك اهتمام كبير في السنوات الأخيرة على طريقة الترجيح الأعظـم لكونهـا أكفأ وأفضل طريقة لتقـدير المعلمـات، وفـي عـام (1996) قـدم الباحثـان Lehr و Lii الخوارزمية الأساسية للتـرجيح الأعظـم (Maximum Likelihood Estimation) التـي تعطي تقديرات ثابتة وكفوءة للمعلمات، إن تقدير العينات الكبيرة يمكن أن تكون مكلفة أو حتى مستحيلة ولهذا طور كل من Xu و Vogl في عام (2000) طريقة الترجيح الأعظم وتمكنـا من تقدير معلمات هذه العينات.

وهناك عدة بحوث تختص بتقدير المعلمات باستخدام توزيع بولتزمان الذي يمثل تركيـب سلسلة الـ RNA وفيما يلي بعض من هذه الأعمال و البحوث:

وصف Pond وآخرون في عام (2006) طريقة أساسها دالة الإمكان لإختيار النموذج التطوري باستخدام الخوارزمية الجينية (GA) ليستكشف بسرعة مجموعة كبيرة متحدة لكل نماذج ماركوف القابلة للانعكاس مع عدد ثابت لإحلال النسب. وحققوا أيضاً فائدة لعدة مقاييس مسافة للمقارنة والتغاير للنماذج التطورية المستنتجة [10].

وفي العام نفسه وصف Ding وآخرون إجراء بديل لتمثيل تركيب الحامض الرايبـي المراسل mRNA، الذي اختبرت فيه التراكيب مـن مجموعـة بولتزمـان (Boltzmann) المرجحة لتراكيب سلسلة الـ RNA الثانوية المتجمعة، بالاعتماد على عينة عشوائية مـن طول طبيعي للإنسان وأخذ الـ mRNA منه، واظهروا خصائص المستوى العنقودي بأنهـا قابلة للإنتاج بشكل إحصائي في مقارنة بين هيكلة mRNA و RNA [8].

كما قدم Andronescu وآخرون في العام نفسه طريقة قيـد الجيـل (CG)، وهـي الطريقة الحسابية الأولى لتقدير معلمة الطاقة الحرة لسلسلة الـ RNA التي يجب أن تـدرب بشكل كفوء على هيكلية المجموعات الكبيرة فضلاً عن بيانات الديناميكية الحرارية. وبينوا أن طريقة قيد الجيل (Constraint Generation) توظف مخططاً تكرارياً مبتكراً، إذ تم حساب الطاقة أولاً كحل لمسالة الأمثلية، ثم استخدموا معلمات الطاقة المحسوبة لتجديد القيـود لدالـة الأمثلية، لكي يحسنوا أمثلية معلمات الطاقة في التكرار القادم [3].

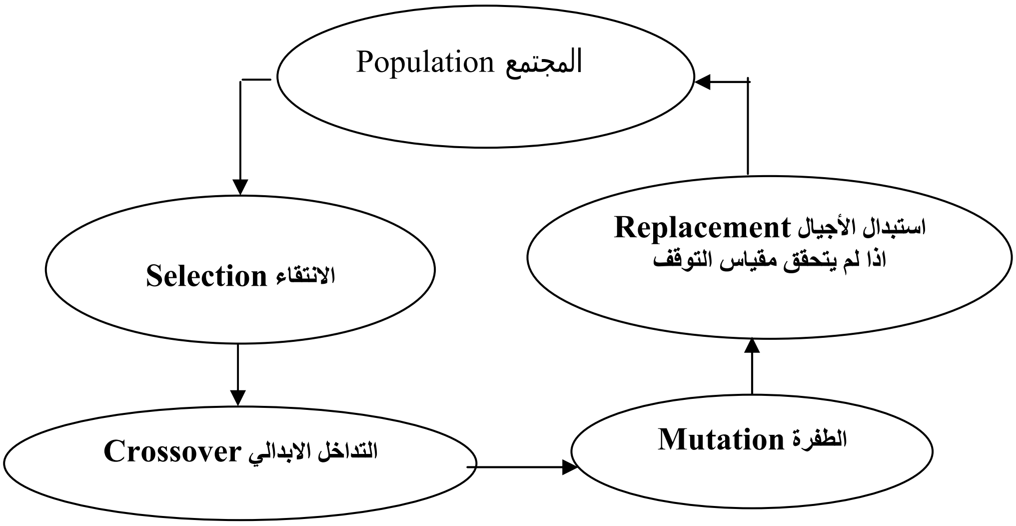
وقد استعمل الباحثان Chan و Ding في عام (2008) أسلوب العينات وجمعوا عناقيد لمجموعة بولتزمان (Boltzmann) لتركيب RNA الثانوي، للتحقق فـي هيئـة المجموعـة المعروضة للسلاسل الحيوية التي ميزت من تعديلاتهم العشوائية، وقاموا بتحليل أساس تركيب الـ MicroRNA المتمثل بـ (miRNA) لـ 9 مجموعات مميزة [4].

**في هذا البحث تم تقدير معلمة توزيع بولتزمـان المتمثـل بمعلمـة RNA باسـتخدام الخوارزمية الجينية** وذلك بـ :

1. تعظيم دالة الترجيح لتوزيع بولتزمان باستخدام الخوارزمية الجينية.
2. تقليل قيد الجيل لتوزيع بولتزمان باستخدام الخوارزمية الجينية.

## 2.الخوارزميـ ـــة الجينيـ ـــة: Genetic Algorithm

تتضمن الخوارزمية الجينية عدداً من الخطوات الأساسية، هذه الخطوات ثابتة لمختلـف المسائل ولكل التطبيقات ويكون الاختلاف في صياغة كل خطوة من الخطوات وتطبيقها حسب المسألة أو مجال تطبيقها، إن خطوات هذه الخوارزمية تكون مترابطة بعضها مـع الـبعض الآخر، ولا يمكن تطبيق هذه الخوارزمية على أية مسألة ما لم تطبق جميع هذه الخطوات وإلاّ تفقد الخوارزمية الجينية قيمتها وفائـدتها فـي إيجـاد أو تحـسين الحـل. [1] [12] [13] والشكل(1) يوضح المخطط العام للخوارزمية الجينية:




**شكل (1) يمثل خطوات الخوارزمية الجينية**

## خطوات عمل الخوارزمية الجينية:

## Working Steps in Genetic Algorithm

**١-البداية Start:** توليد مجتمع عشوائي من الكروموسومات، أي بمعنى إيجاد حلول مناسبة للمسألة.

**٢-اللياقة Fitness:** هي تحويل دالة الهدف (Objective Function) إلى دالة مناسبة للحل في الخوارزمية الجينية.

**٣-مجتمع جديد New Population:** توليد جيل جديد بتكرار الخطوات الآتية إلى أن يكتمل الجيل، وتتضمن:

- الاختيار Selection: يتم اختيار اثنين من الكروموسومات والدين ( Parents Chromosomes) من المجتمع الابتدائي استنادا إلى دالة اللياقة (أفضل القيم التي لها فرص اكبر للاختيار).
- التداخل الابدالي Crossover: إجراء إحدى عمليات التعابر للحصول على الذرية ( Offspring) ويكون بين كروموسومين.
- الطفرة Mutation: مع احتمالية وجود الطفرة يتم عمل الطفرة للسلف الجديد بموقع معين في الكروموسوم، وتجري بين الجينات في الكروموسوم الواحد.
- الاستبدال (Replacement): عملية وضع السلف الجديد المتكون في الجيل الجديد للحلول محل المجتمع الابتدائي.
- الاختبار Test: عند تحقق شرط التوقف، فإن الخوارزمية الجينية تتوقف وتعيد الحل الجيد من آخر جيل متكون.
- الدورة Loop: يتم الرجوع إلى الخطوة 2.

سإن كل تكرار لهذه العملية يسمى بالجيل ( Generation)، وبعد نهاية التنفيذ يقوم الباحث بتقديم تقرير عن الحقائق التي تم التوصل إليها.

## 3. الأحماض النووية في الخلية: Nucleic of the Cell

توجد الأحماض النووية في الخلية"، إذ أن الخلية هي وحدة البناء والوظيفة في الكائن الحي وهي على نوعين:الحامض النووي الرايبي (Ribo Nucleic Acid) ويرمز له اختصاراً RNA والحامض النووي الرايبي منقوص الأوكسجين ( Deoxyribo Nucleic Acid) ويرمز له DNA [11].

إن موضوع دراستنا يتناول الحامض النووي الرايبوسومي (RNA) وهو يوجد في كل من النواة والسايتوبلازم، حيث يوجد في النوية وفي الرايبوسومات وفي تراكيب أخرى.

ويمكن تعريف (السايتوبلازم) بأنه جزء المادة الحية للخلية الذي يقع خارج النواة. وتعرف (النواة) على أنها أكبر عضية متميزة داخل الخلية وللنواة أهمية كبيرة في نقل الصفات الوراثية وفي النشاط الأيضي للخلية [9].

تتكون جزيئة الحامض النووي الرايبي (RNA) من شريط واحد، إذ أن النيوكليوتيدات الداخلة في تركيب الـ RNA تتألف من ثلاث جزيئات بسيطة مرتبطة بعضها ببعض مباشرة وهي:

١- قاعدة نتروجينية: وهي مركب حلقي يحتوي على النيتروجين فضلاً عن الكاربون والهيدروجين والأوكسجين عدا الأدنين الذي لا يحتوي بدوره على الأوكسجين، ويوجد نوعان منها:

أ-البريميدينات (Pyrimidines): وتتكون من حلقة واحدة وتشمل القواعد الاتية:

١-السايتوسين (Cytosine) ويرمز لها (C).

٢-اليوراسيل (Uracil) ويرمز لها (U).

ب- البيورينات (Purines): وتتكون من حلقتين وتشمل:

١- الأدنين (Adenine) ويرمز لها (A).

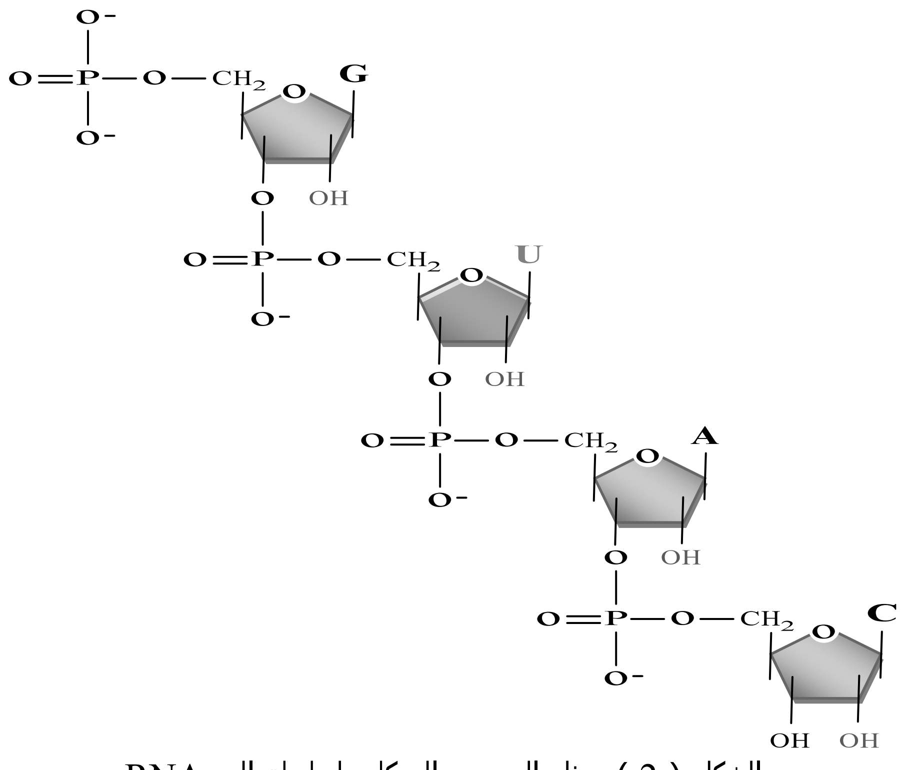
٢- الكوانين(Guanine) ويرمز لها (G).

فالأدنين يرتبط دائماً مع اليوراسيل بآصرتين هيدروجينيتين A=U والسايتوسين يرتبط دائماً مع الكوانين بثلاث أواصر هيدروجينية C ≡ G.

٢- سكر خماسي الكاربون وهو الرايبوز وصيغته الجزيئية هي ($C_5H_{10}O_5$).

٣- حامض الفسفوريك.

والشكل الآتي يبين العرض الهيكلي لسلسلة الـ RNA




الشكل ( 2 ) يمثل العرض الهيكلي لسلسلة الـ RNA

وفي حالات نادرة قد يكون الحامض النووي الرايبوسومي RNA هو المادة الوراثية كما في بعض الرواشح (الفيروسات)، ولكن أهميته أعظم وأعم في الأحياء بسبب الدور الذي يقوم به في عملية بناء البروتين بما في ذلك الأنزيمات.

ويمكن تمييز ثلاثة أنواع من الـ RNA وكلها تصنع في النواة وتنتقل إلى السايتوبلازم لتشارك في صنع البروتين بشكل خاص وهي:

١- الحامض الرايبي المراسل Messenger RNA (mRNA): ويقوم بنقل المعلومات الوراثية من DNA النواة إلى مناطق نشاطه في السايتوبلازم.

٢- الحامض الرايبي الرايبوسومي Ribosomal RNA (rRNA): ويدخل في تركيب الرايبوسومات إذ يشترك مع البروتين في بنائها.

٣- الحامض النووي الرايبي الناقل Transfer RNA (tRNA): ويوجد في السايتوبلازم ويقوم بنقل الأحماض الأمينية إلى الرايبوسومات.

## 4. توزيع بولتزمان: Boltzmann Distribution

تعرف دالة كثافة الاحتمال لتوزيع بولتزمان بالشكل الآتي: [3]

$$P(x;\theta)=\begin{cases}\exp\dfrac{-\Delta G(x;\theta)}{kT} & if \quad \Delta G(x;\theta)>0\\ 1 & if \quad \Delta G(x;\theta)\le 0\end{cases} \qquad (1)$$

إذ أن:

$\Delta G$ : التغير في الطاقة، وتعرف بالصيغة الآتية:

$$\Delta G=\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\,\frac{y!}{i!\,(y-i)!}\,(y-i)^{x}\,\theta$$

k: ثابت بولتزمان.

T: درجة الحرارة المطلقة.

## 5. تقدير معلمة توزيع بولتزمان باستخدام الخوارزمية الجينية:

## The Estimate Parameter of the Boltzmann Distribution Using Genetic Algorithm

سوف يتم في هذا البحث اقتراح خوارزميتين لتقدير معلمة توزيـع بولتزمـان باسـتخدام الخوارزمية الجينية التي بدورها سوف تقوم بالمهام التالية:

أ) تعظيم دالة الترجيح لتوزيع بولتزمان.

ب) تقليل قيد الجيل لتوزيع بولتزمان.

### 5.1. تعظيم دالة الترجيح لتوزيع بولتزمان باستخدام الخوارزمية الجينية:

### Maximizing the Likelihood Function of Boltzmann Distribution by Using Genetic Algorithm

سوف يتم إيجاد تقدير المعلمة θ لتوزيع بولتزمان الشرطي باستخدام طريقة الترجيح الأعظم، وهنا سنعرف سلسلة الـ (RNA ) بـ (x)، ونعرف الاحتمالية لتركيب سلسلة الـ (RNA) بـ (y )، ونستخدم النموذج الخطي- اللوغاريتمي الشرطي لتوزيع بولتزمان ( Boltzmann Distribution) ثم نقوم بتعظيم دالة الترجيح لتوزيع بولتزمان بإستخدام الخوارزمية الجينية.

### 5.2 اشتقاق الـ MLE لتوزيع بولتزمان

فيما يأتي الاشتقاق الكامل لتقدير الترجيح الأعظـم (M.L.E. ) لتوزيـع بولتزمـان الشرطي، إذ أن دالة كثافة الأحتمال لتوزيع بولتزمان الشرطي (Conditional Boltzmann Distribution) تعرف كالآتي:[3]

$$p(y/x,\theta)=\frac{1}{z(x,\theta)}\exp(\frac{-1}{kT}\Delta G(x,y,\theta) \qquad (٢)$$

إذ أن ΔG تعرف بالصيغة الآتية:

$$\Delta G=\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta \qquad (٣)$$

وأن الدالة الجزئية z(x,θ) تمثل بـ:

$$z(x,\theta) = \sum_{y} \exp(\frac{-1}{kT} \Delta G(x, y, \theta)) \qquad (٤)$$

$$p(y/x,\theta) = \frac{\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}{\sum_{y}\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)} \qquad (٥)$$

ولإيجاد مقدر الترجيح الأعظم لتوزيع بولتزمان الشرطي نتبع الخطوات الآتية:

أولا": نجد دالة الإمكان L(θ) كالآتي:

$$L(\theta) = \prod_{r=1}^{n} p(y/x,\theta)$$

إذ أن: $\prod_{r=1}^{n} = \sum_{m=1}^{16}\sum_{m=1}^{16}$

$$= \prod_{r=1}^{n} \frac{\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}{\sum_{y}\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)} \qquad (٦)$$

$$= (\frac{\exp\sum_{r=1}^{n}(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}{\prod_{r=1}^{n}\sum_{y}\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}) \qquad (٧)$$

ثانيا": نأخذ اللوغاريتم لدالة L(θ) لتصبح المعادلة كالآتي:

$$\ln L(\theta) = \log(\frac{\exp\sum_{r=1}^{n}(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}{\prod_{r=1}^{n}\sum_{y}\exp(\frac{-1}{kT}\sum_{x}\frac{1}{y!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y!}{i!(y-i)!}(y-i)^{x}\theta)}) \qquad (٨)$$

وبتبسيط المعادلة أعلاه نحصل على ما يلي:

$$\log L(\theta) = -\sum_{x=1}^{4}\sum_{y=1}^{4} (\ln \sum_{m=1}^{16}\sum_{m=1}^{16} \exp(\frac{-1}{kT}\sum_{x} \frac{1}{y(m)!}\sum_{i=0}^{y-1}(-1)^{i} \frac{y(m)!}{i!(y(m)-i)!}(y(m)-i)^{x(m)}\theta) - \sum_{m=1}^{16}\sum_{m=1}^{16}(\frac{-1}{kT}\sum_{x}\frac{1}{y(m)!}\sum_{i=0}^{y-1}(-1)^{i}\frac{y(m)!}{i!(y(m)-i)!}(y(m)-i)^{x(m)}\theta)) \quad (٩)$$

إن دالة كثافة الأحتمال لتوزيع بولتزمـان الـشرطي (Conditional Boltzmann Distribution) تعرف كالآتي: [3].

$$p(y/x,\theta) = \frac{1}{z(x,\theta)}\exp(\frac{-1}{kT}\Delta G(x,y,\theta)$$

## 5.3 الخوارزمية الجينية المقترحة (1):
## Proposed Genetic Algorithm (1)

### إيجاد قيمة المقدر الذي يعظم دالة الترجيح لتوزيع بولتزمان
### Finding the Value of the Estimator which Maximizes the Likelihood Function for Boltzmann Distribution

تم استخدام دالة الترجيح الأعظم لتوزيع بولتزمان في الخطوات المقترحة للخوارزميـة الجينية، والخطوات موضحة في أدناه :

**الخطوة الأولى:- البيانات الأولية (Initial Data):** وهي قراءة لمجموعة من المتغيرات التي استخدمت في الخوارزمية:

- العداد m: وهو يمثل الطول لكل من سلسلتي x و y .
- x(m): يمثل متجه لأحد النيوكليوتيدات الداخلة في تركيب الـ RNA ويمثـل القواعد النتروجينية: (أدنين A، يوراسيل U، سايتوسين C، كـوانين G) وكـل عنصر مكرر أربع مرات، إذ تم تشفير هذه التراكيب ورمزنا للأدنـين بـالرمز (000)، ولليوراسيل بالرمز (001)، وللـسايتوسين بـالرمز (111)، والكـوانين بالرمز (010).

- y(m): يمثل متجه التراكيب الناتجة مـن تـداخل هـذه القواعـد النتروجينيـة وللتكرارات أعلاه (فقط التراكيب المقبولة التي فيهـا يـرتبط الأدنـين A مـع اليوراسيل U بآصرتين هيدروجينيتين A=U ويرتبط السايتوسين C مع الكوانين G بثلاث أواصر هيدروجينية C≡G)، وبهذا سوف يصبح تركيب الـ AU هـو (000001)، وتركيب الـ CG هو (111010)، وتركيـب الــ UA هـو (001000)، وتركيب الـ GC هو (010111).
- k: يمثل ثابت الغاز.
- T: تمثل درجة الحرارة المطلقة.

**الخطوة الثانية:- إنشاء الجيل الابتدائي (Initial Generation):** الجيل الابتدائي فـي هذه الخوارزمية يتكون من كروموسومين احدهما يمثل القواعد النتروجينية والآخر يمثـل تراكيب هذه القواعد النتروجينية ، وقد تم وضع القيم الحقيقية في جينات الكروموسـومين أي أن التشفير كان تشفيرا حقيقيا للكروموسومات.

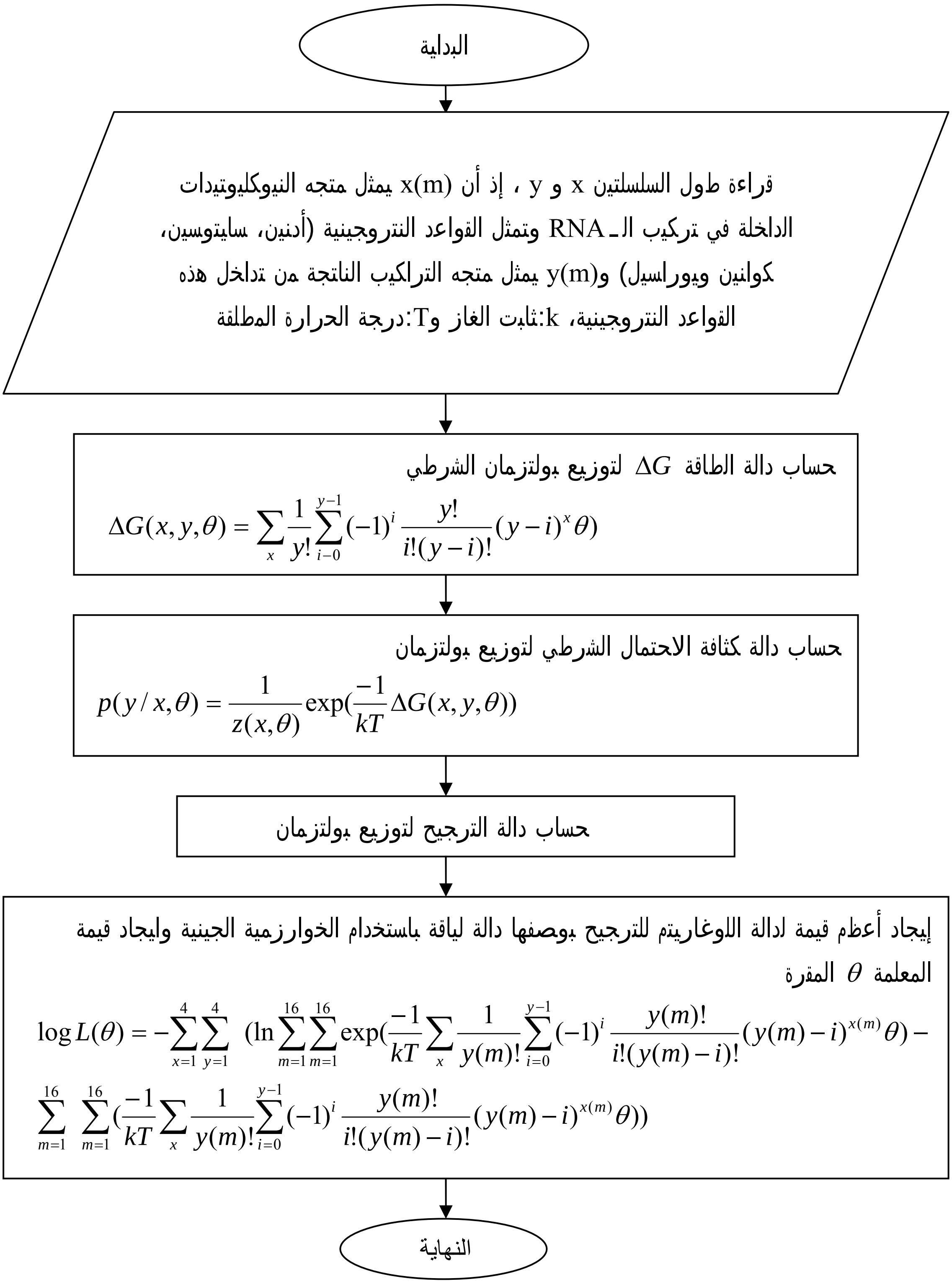
**الخطوة الثالثة:- قيمة الجودة ( Fitness Value):** إن قيمة الجودة في هذه الخوارزمية تمثل مقدر الترجيح الأعظم لتوزيع بولتزمان، إذ تم استخدام اللوغاريتم لأنه الأساس فـي استخراج المقدرات التي تجعل المقدر أعظم ما يمكن.

**الخطوة الرابعة:-** تم استخدام الانتقاء من النوعين الآتيين **(**Uniform ،Roulette**).**

**الخطوة الخامـسة:-** تم استخدام التعابر من الأنواع الآتية **(**Scattered ،Intermediate ،Single Point، Heuristic)

**الخطوة السادسة:-** تم استخدام الطفرة من الأنواع الآتية **(**Uniform، Gaussian).
وتم التحديد المسبق لعدد الأجيال في توقف الخوارزمية الجينية.

والشكل الآتي يمثل المخطط الانسيابي للخوارزمية الجينية المقترحة:




شكل (3) يمثل المخطط الانسيابي لتقدير قيمة المعلمة θ
الذي يعظم دالة الترجيح لتوزيع بولتزمان

تم تصميم برنامج بلغة (MATLAB) ثم تم ايجاد قيمة دالة الجودة (أعظم قيمـة لدالة الترجيح) ومن خلاله تم تحديد المعلمة التي ستحقق القيمة العظمى للدالة وتم الاعتماد على التحديد المسبق لعدد الأجيال لبيان مدى التقرب من الحل الأمثل وقمنا باختيار عـدة أنواع لكل من (الانتقاء ، التداخل الابدالي والطفرة) وكذلك عدد مـرات توليـد الأجيـال وملاحظة النتائج بعد كل جيل إذ كانت أفضل النتائج لدى توليد الأجيال مـابين 25 إلـى 1500 جيل ولقد حققـت هذه الخوارزمية أعظم قيمـة لمقدر الترجيح لتوزيع بولتزمـان عنـد القيمة (**-94.1614**) وقيمة θ هي (0.1457) والتي ظهرت عند الجيل (1387) إذ كان:

١. الانتقاء من نوع (Roulette).
٢. التداخل الابدالي من نوع (Intermediate).
٣. الطفرة من نوع (Uniform).

كما موضح في الجدول (1) .

| | عدد الأجيال | نوع الانتقاء | نوع التداخل الابدالي (التعابر) | نوع الطفرة | قيمة θ | أعظم قيمة لمقدر M.L.E |
|---|---|---|---|---|---|---|
| 1. | 100 | Roulette | Intermediate | Uniform | -21.0133 | -110.018 |
| 2. | 77 | Roulette | Intermediate | Uniform | -17.1375 | -102.0289 |
| 3. | ٨٠ | Uniform | Intermediate | Gaussian | -8.0851 | -95.5772 |
| 4. | ٢٩ | Uniform | Scattered | Gaussian | -4.1539 | -94.7458 |
| 5. | 100 | Uniform | Scattered | Gaussian | -12.9348 | -97.6835 |
| 6. | ٢٥ | Uniform | Single Point | Uniform | 0.0229 | -94.1751 |
| 7. | ٥٢ | Uniform | Intermediate | Uniform | 0.044 | -94.1727 |
| 8. | ١٠٠ | Roulette | Scattered | Gaussian | -23.5084 | -117.5638 |
| 9. | ٧٣ | Roulette | Intermediate | Uniform | -18.9013 | -105.1023 |
| 10. | ١٠٠ | Roulette | Single Point | Uniform | 0.0084 | -94.1767 |
| 11. | 171 | Uniform | Scattered | Gaussian | -٧,٧٧٦٧ | -٩٥,٤٩٤٦ |
| 12. | ١٠٠ | Uniform | Single Point | Gaussian | -13.0772 | -97.7796 |
| 13. | 100 | Uniform | Scattered | Gaussian | -15.6705 | -100.1023 |
| 14. | ١٠٠ | Roulette | Heuristic | Uniform | -29.8607 | -142.7702 |
| 15. | ١٠٠ | Stochastic Uniform | Scattered | Gaussian | -20.5965 | -108.9377 |
| 16. | ٨٤ | Roulette | Intermediate | Uniform | -18.2274 | -103.8222 |
| 17. | ٥٩ | Uniform | Heuristic | Gaussian | -17.3304 | -102.3224 |
| 18. | 99 | Roulette | Intermediate | Uniform | -19.8354 | -107.105 |
| 19. | ٣٥ | Uniform | Scattered | Gaussian | -5.3125 | -94.9497 |
| 20. | ٤٩ | Uniform | Single Point | Uniform | 0.1016 | -94.1663 |
| 21. | ١٣٦ | Roulette | Scattered | Uniform | ٠,٠٨١٥ | -٩٤,١٦٨٥ |
| 22. | ٢٣٤ | Uniform | Scattered | Gaussian | -1.8355 | -٩٤,٤٠٢٦ |
| 23. | ٥٦٧ | Roulette | Intermediate | Uniform | ٠,٠٠٦٠ | -٩٤,١٧٧٠ |
| 24. | ٧٨٤ | Roulette | Single Point | Gaussian | ٠,٠٣٢٦ | -٩٤,١٧٤٠ |
| 25. | ٩٢٣ | Uniform | Scattered | Uniform | ٠,٠٤٣٥ | -٩٤,١٧٢٨ |
| 26. | ٤١٩ | Roulette | Single Point | Gaussian | -45.8077 | -٢١٦,٤٠٩٣ |
| 27. | ١٠٩٧ | Roulette | Heuristic | Uniform | -2.1757 | -٩٤,٤٤٨٥ |
| 28. | ١٢٦٥ | Uniform | Scattered | Gaussian | -5.7313 | -٩٥,٠٣٠٤ |
| 29. | **1387** | **Roulette** | **Intermediate** | **Uniform** | **٠,١٤٥٧** | **-٩٤,١٦١٤** |
| 30. | ١٥٠٠ | Uniform | Single Point | Gaussian | -4.2841 | -٩٤,٧٦١٤ |

جدول (1) يمثل تقدير المعلمة θ بأسلوب تعظيم دالة الترجيح لتوزيع بولتزمان

## 5.4 تقليل قيد الجيل لتوزيع بولتزمان باستخدام الخوارزمية الجينية:
## Minimizing Constraint Generation (CG) of Boltzmann Distribution Using Genetic Algorithim

الطريقة البديلة لطريقة الترجيح في تقدير المعلمة هي طريقة قيد الجيل والتي اقترحت من قبل Andronescu وآخرون في عام (٢٠٠٦) والطريقة تتلخص بتقدير قيمـة θ لتوزيـع بولتزمان بحل نظام القيود الموضح في المعادلة الآتية:[10]

$$\Delta G(x,y_x,\theta) \leq \Delta G(x,y,\theta) \qquad (10)$$

لكل $(x,y_x)\in S$ و $y\in Y_x\backslash\{y_x\}$

إذ أن:

S: تمثل مجموعة بيانات تتكون من مجموعة أزواج $(x,y_x)$.

$y_x$: تمثل مجموعة كل التراكيب الثانوية للمتسلسلة x.

هذه القيود تضمن أن لكل سلسلة x جميع تراكيبها الثانوية y غير المثلـى والتـي يرمز لها بـ $(y_x)$ لها أعلى طاقة حرة دنيا. (من خلال ذلك نفرض انه لا يوجد تركيـب آخر له نفس الطاقة الحـرة الـدنيا (Minimum Free Energy) (MFE) كالتركيـب المعرف لدينا).

قد يحدث أن هذا النظام من القيود يكون غير قابل للتطبيق، أي انه لا يوجـد حـل للمعلمة (θ) الموجودة في توزيع بولتزمان والتي تعرف كل القيود بشكل آني. وللتعامـل مع التطبيقات غير الممكنة، سنضيف المتغيرات البطيئة $\delta_{x,y} \geq ٠$ إلى القيود، والتي سوف يتم تقليل قيمتها، وهذا يؤدي إلى استرخاء (Relaxation) القيود (أي تحسين قيمتها) كما موضح في المعادلة الآتية:

$$\Delta G(x,y_x,\theta) \leq \Delta G(x,y,\theta) + \delta_{x,y} \qquad (11)$$

وعلى اعتبار أن دالة الطاقة ΔG معرفة سابقاً في المعادلة (٣)، ويمكن تحويل المتراجحة أعلاه إلى معادلة معرفة كالآتي:

$$\Delta G(x,y_x,\theta) - \Delta G(x,y,\theta) = \delta_{x,y} \qquad (12)$$

وبالتالي يجب أن تتحقق دالة قيد الجيل والممثلة بالمعادلة الآتية:

$$\text{minimize } \|\delta\|^2 \qquad (13)$$

مع تحقق القيد:

$$\delta \geq 0 \qquad (14)$$

تم استخدام المعادلة (13) مع القيد (14) كدالة هدف للخوارزمية الجينية المقترحة التالية والتي يتم فيها تقليل دالة الهدف مع تحقق قيدها.

## 5.5 الخوارزمية الجينية المقترحة (2):
## Proposed Genetic Algorithm (2)

### إيجاد قيمة المقدر الذي يقلل قيد الجيل لتوزيع بولتزمان
### Finding the Value of the Estimator which Minimizes the Constraint Generation for Boltzmann Distribution

استخدمت دالة قيد الجيل (CG) لتوزيع بولتزمان في الخطوات المقترحة للخوارزمية الجينية، والخطوات موضحة في أدناه:

**الخطوة الأولى والخطوة الثانية كما في مثيلتيهما في الخوارزمية السابقة.**

**الخطوة الثالثة:- قيمة الجودة ( Fitness Value):** إن قيمة الجودة في هذه الخوارزمية ممثلة بالمعادلة (٣,١٤) إذ استُخدم الرمز z ليمثل ناتج طرح دوال الطاقة في الحالتين: عند تكرار عدد التراكيب $\Delta G(x,y_x,\theta)$، وعند ثبوت عدد التراكيب التي رمز لها بـ $\Delta G(x,y,\theta)$ ، أي أن:

$$z = \Delta G(x,y_x,\theta) - \Delta G(x,y,\theta)$$

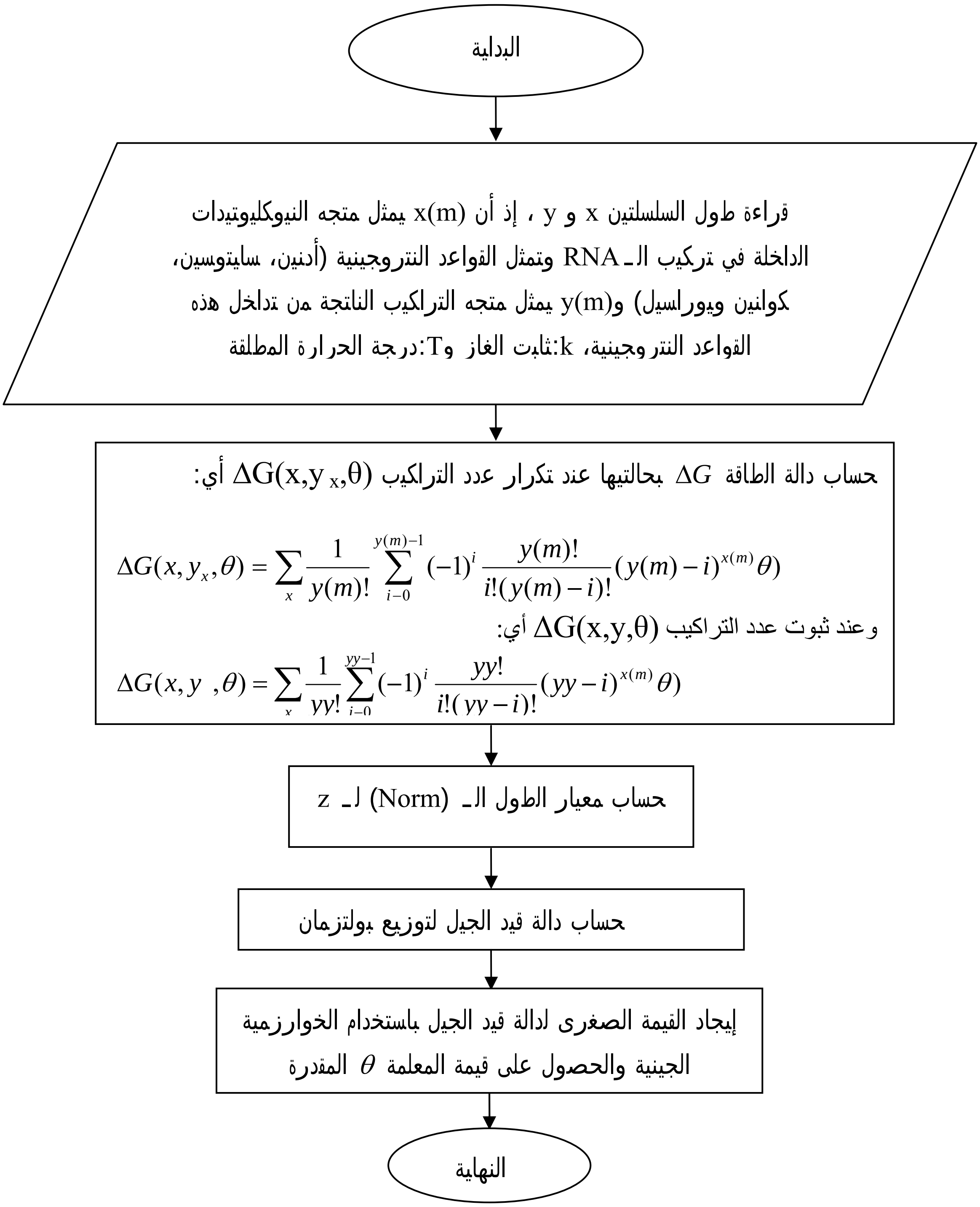
وفي هذه الحالة يجب تحقق الشرط التالي وهو أن z أكبر أو تساوي الصفر، عندها سوف يتم حساب معيار الطول (Norm) لـ z التي تمثل المعادلة (٣,١٤) بعدها سوف يتم تقليل قيمتها.

**الخطوة الرابعة:-** تم استخدام الانتقاء من النوعين الآتيين **(**Roulette، Uniform**).**

**الخطوة الخامـسة:-** تم استخدام التعابر من الأنواع الآتية **(**Scattered، Intermediate، Single Point، Two Point، Heuristic)

**الخطوة السادسة:-** تم استخدام الطفرات الآتية **(**Uniform، Gaussian).
وتم التحديد المسبق لعدد الأجيال في توقف الخوارزمية الجينية.

والشكل التالي يمثل المخطط الانسيابي للخوارزمية الجينية المقترحة:




شكل (4) يمثل المخطط الانسيابي لإيجاد قيمة المعلمة θ
الذي يقلل دالة قيد الجيل لتوزيع بولتزمان

بعد أن تم ايجاد قيمة الجودة (أقل قيمة لدالة قيد الجيل)، وذلك بتصميم برنـامج بلغة MATLAB ومن خلاله تم تحديد المعلمة التي ستحقق القيمة الدنيا للدالة وتم اختيار عدة أنواع لكل من (الانتقاء، التداخل الابدالي والطفرة) وكذلك عدد مرات توليد الأجيـال وملاحظة النتائج بعد كل جيل إذ كانت أفضل النتائج لدى توليد الأجيال مـابين ٢٥ إلـى ١٥٠٠ جيل وتم الاعتماد على التحديد المسبق لعدد الأجيال لبيان مدى التقرب من الحـل الأمثل ولقد حققت هذه الخوارزمية أقل قيمة لدالة قيد الجيل لتوزيع بولتزمان عند القيمـة ($0,951039101*10^7$) وقيمة θ هي (4.4066-) والتي ظهرت عنـد الجيـل (٨٠) إذ كان:

١. الانتقاء من نوع (Uniform).
٢. التداخل الابدالي من نوع (Intermediate).
٣. الطفرة من نوع (Gaussian).

كما موضح في الجدول (٢).

جدول (2) يمثل تقدير المعلمة θ بأسلوب تقليل قيد الجيل CG لتوزيع بولتزمان

| | عدد الأجيال | نوع الانتقاء | نوع التداخل الابتدائي (التقاطع) | نوع الطفرة | قيمة θ | أقل قيمة لمقدر قيد الجيل (CG) |
|---|---|---|---|---|---|---|
| 27. | ١٢١٥ | Roulette | Scattered | Uniform | -5.2872 | $-0.13691*10^{1}$ |
| 28. | ١٣٦٥ | Uniform | Intermediate | Gaussian | -1.4349 | $-0.100842706*10^{7}$ |
| 29. | ١٤٧٧ | Roulette | Scattered | Uniform | 0.9489 | $-0.19399*10^{1}$ |
| 30. | ١٥٠٠ | Uniform | Single Point | Uniform | ٠,٩٢٠٧ | $-0.18263*10^{1}$ |
| 1. | ٦٤ | Roulette | Two Point | Uniform | ٠,٩٢٦٧ | -٠,١٨٥٠٢*$10^{1}$ |
| 2. | ١٠٠ | Stochastic Uniform | Scattered | Gaussian | -21.6482 | -٠,٢٢٩٥٢*$10^{1}$ |
| 3. | ٤٩ | Uniform | Single Point | Uniform | ٠,٩٩٤٤ | -٠,٢١٣٠٣*$10^{1}$ |
| 4. | ٢٩ | Uniform | Scattered | Gaussian | -4.6912 | -٠,١٠٧٧٨*$10^{1}$ |
| 5. | ٩٩ | Roulette | Intermediate | Uniform | ٠,٩٨٩٥ | -٠,٢١٠٩٦*$10^{1}$ |
| 6. | ١٠٠ | Roulette | Intermediate | Uniform | ٠,٩٧٧٦ | -٠,٢٠٥٨٩*$10^{1}$ |
| 7. | ١٠٠ | Uniform | Scattered | Gaussian | -10.9584 | -٠,٥٨٨١٤*$10^{1}$ |
| 8. | ١٢٣ | Uniform | Scattered | Gaussian | -3.4385 | -٠,٥٧٩٠٨١٢٦٥*$10^{7}$ |
| **9.** | **٨٠** | **Uniform** | **Intermediate** | **Gaussian** | **-4.4066** | **-٠,٩٥١٠٣٩١٠١*$10^{7}$** |
| 10. | ١٠٠ | Uniform | Single Point | Gaussian | -12.4559 | -٠,٧٥٩٨٧*$10^{1}$ |
| 11. | 52 | Uniform | Intermediate | Uniform | ٠,٩٦٠٩ | -0.١٩٨٩٢*$10^{1}$ |
| 12. | ٧٣ | Roulette | Intermediate | Uniform | ٠,٨٥٦٩ | -٠,١٥٨٢٠*$10^{1}$ |
| 13. | 25 | Uniform | Single Point | Uniform | ٠,٩٩١٦ | -٠,٢١١٨٤*$10^{1}$ |
| 14. | 100 | Uniform | Scattered | Gaussian | -6.5909 | -٠,٢١٢٧٦*$10^{1}$ |
| 15. | 84 | Roulette | Intermediate | Uniform | ٠,٨٨٣٤ | -٠,١٦٨١٥*$10^{1}$ |
| 16. | ٥٩ | Uniform | Heuristic | Gaussian | -26.4713 | -٠,٣٤٣١٩*$10^{1}$ |
| 17. | ٣٥٠ | Uniform | Scattered | Gaussian | -1.7094 | $-0.143105685*10^{7}$ |
| 18 | 100 | Roulette | Heuristic | Uniform | ١,٣٨٥٨ | $-0.41378*10^{1}$ |
| 19. | 77 | Roulette | Intermediate | Uniform | ٠,٩٦٢٥ | -٠,١٩٩٦١*$10^{1}$ |
| 20. | 100 | Roulette | Scattered | Gaussian | -21.5456 | $-0.22736*10^{1}$ |
| 21. | ١٧٨ | Uniform | Heuristic | Gaussian | -50.4029 | $-0.12442*10^{1}$ |
| 22. | ٢٣٢ | Roulette | Intermediate | Gaussian | -7.8907 | -٠,٣٠٤٩٤٣*$10^{1}$ |
| 23. | ٤٨٩ | Uniform | Scattered | Uniform | ٠,٩٨٨٩ | $-0.21066*10^{1}$ |
| 24. | ٦٧٨ | Uniform | Intermediate | Gaussian | -0.5562 | $-0.15151389*10^{6}$ |
| 25. | ٨١ | Uniform | Intermediate | Gaussian | -8.5679 | $-0.35953*10^{1}$ |
| 26. | ١٠١٠ | Uniform | Scattered | Gaussian | -1.8179 | $-0.16184835810^{7}$ |

## 6. الاستنتاجات:

إيجاد قيمة المقدر الذي يعظم دالة الترجيح لتوزيع بولتزمـان باسـتخدام الخوارزميـة الجينية المقترحة الأولى استُنتج ما يأتي:

أعظم قيمة لمقدر الترجيح لتوزيع بولتزمان ظهرت:

١. عند القيمة (-٩٤,١٦١٤) حيث كانت قيمة θ هي (٠,١٤٥٧).
٢. عند الجيل (١٣٨٧).
٣. الانتقاء من نوع (Roulette).
٤. كذلك التداخل الابدالي من نوع (Intermediate).
٥. الطفرة من نوع (Uniform).

إيجاد قيمة المقدر الذي يقلل قيد الجيل لتوزيع بولتزمان باسـتخدام الخوارزميـة الجينيـة المقترحة الثانية استُنتج ما يأتي:

أقل قيمة لقيد الجيل لتوزيع بولتزمان ظهرت:

١. عند القيمة (-٠,٩٥١٠٣٩١٠١*$10^7$) حيث كانت قيمة θ هي (-4.4066).
٢. عند الجيل (٨٠).
٣. الانتقاء من نوع (Uniform).
٤. التداخل الابدالي من نوع (Intermediate).
٥. الطفرة من نوع (Gaussian).

استنتجنا مما سبق بأن الخوارزمية الجينية كانت الأسلوب الأفضل لتحقيق أعظـم قيمـة لدالة الترجيح والحصول على أمثل قيمة للمعلمة θ وكذلك حققت أدنى قيمة لدالة قيـد الجيـل والحصول أيضاً على أمثل قيمة للمعلمة θ.

## 7. المصادر:

1- ثابت، همسة معن. (٢٠٠٥)، "**بعض تطبيقات الخوارزمية الجينيـة فـي حـل مـسائل الأمثلية**"، رسالة ماجستير، جامعة الوصل، كلية علوم الحاسبات والرياضيات، قسم الإحصاء.

2- ذنون، باسل يونس (١٩٨٨)، "**الإحتمالية والمتغيرات العـشوائية**"، مديريـة دار الكتـب والنشر، جامعة الموصل، وزارة التعليم العالي والبحث العلمي.

٣- Andronescu, M.; Condon, A.; Hoos, H. H.; Mathews, D. H. and

Murphy, K., (2006), " **Efficient Parameter Estimation for RNA Secondary Structure Prediction",** © Oxford University Press.

4- Chan, C. Y. and Ding, Y., (2008), " **Boltzmann Ensemble Features of RNA Secondary Structures: a Comparative Analysis of Biological RNA Sequences and Random Shuffles",** C. Y. Chan · Y. Ding Wadsworth Center, New York State Department of Health, Center for Medical Science, 150 New Scotland Avenue, Albany, NY 12208, USA.

5- Devore, J. R., (2000), **"Probability and Statistics"**, Brooks, Cole.

6- Ding, Y. and Lawrence, C. E., (2003), **"A statistical Sampling Algorithm for RNA Secondary Structure Prediction. Nucleic Acids Research"**, 31: pp. 7280–7301.

7- Ding, Y., (2006), **"Statistical and Bayesian Approaches to RNA Secondary Structure Prediction"**, Published by Cold Spring Harbor Laboratory Press, 12: pp. 323–331.

8- Ding, Y.; Chan, C. Y. and Lawrence, C. E., (2006), " **Clustering of RNA Secondary Structures with Application to Messenger RNAs"**, Elsevier Ltd, 0022-2836.

9- Irwin, H. H., (1967), **"Basic Principle of Molecular Genetice"**, Nelson Ltd. London.

10- Pond, K. S. L.; Mannino, F. V. ; Gravenor, M. B.; Muse, S. V. and Frost, S. D. W., (2006), **"Evolutionary Model Selection with a Genetic Algorithm: A Case Study Using Stem RNA"**, Oxford University Press on Behalf of the Society for Molecular Biology and sEvolution. Moscow.

11- William, T. T. and Richared, J. W., (1968), **"General Biology",** 2nd Edition, Van Nostrand Reinhold Company.

12- **"Genetic Algorithms",** the web site at http://www.cs.felk.cvut.cz/~xobitko/ga/main.html Obitko M. (2009)

13- Marczyk, A., (2004), **"Genetic Algorithm and Evolutionary Computation",** April, Vol. 23.

http://www.talkor:gins.org/Faqs/genalg/genalg.html/#introduction.