المجلة العراقية للعلوم الاحصائية (13) 2008
ص ص [152–162]

استخـدام الخـوارزمـية الجينـية
فـي تعظـيم دالـــة الامكـان للـتوزيع الطبيـعي

همسة معن محمد ثابت*

الملخص

تم في هذا البحث تطبيق الخوارزمية الجينية التي هي طريقة بحث اصطناعي لإيجاد القيمة التي تجعل دالة الامكان أعلى ما يمكن (تعظيم دالة الامكان) إذ تم اقتراح خوارزمية لإيجاد القيم التي تعظم دالة الامكان للتوزيع الطبيعي وأدى تطبيقها إلى إيجاد عدد من الحلول بعدد مرات توليد الخوارزمية ومن ضمنها القيمة التي تحقق تعظيم دالة الامكان.

Using the Genetic Algorithm to Maximize the Likelihood Function of Normal Distribution

ABSTRACT

In this research, the genetic algorithm (GA) has been carried out. This application is considered to be a manufacturing treatment to find the value which maximizes the likelihood function . An algorithm has been proposed to find the values which maximize. The likelihood function for normal distribution. The application of this algorithm enables us to find out several solutions, including the value which is responsible for the maximization of the likelihood function and the number of the latter is equal to the times for generating the algorithm.

1- المقدمة Introduction

إن لنظرية التقدير أهمية كبيرة في تطبيقات النظرية الإحصائية في الجوانب العملية، ويمكن تجزئتها الى جزأين متكاملين، الأول يهتم بالبحث عن أفضل تقدير

• مدرس مساعد - جامعة الموصل - كلية الآداب

تاريخ التسلم : 2007/2/1 ـــــــــــــــ تاريخ القبول : 2007/5/6

لمعلمة مجهولة في المجتمع ويسمى التقدير بنقطة (Point of Estimation) في حين إن الجزء الثاني يهتم بالبحث عن أفضل فترة يمكن حصر قيمة المعلمة المجهولة خلالها وهذا ما يسمى بتقدير الفترة (Interval Estimation) (8) وفيما يلي عرض لأهم طرائق تقدير النقطة:

طرائق التقدير Methods of Estimation

هنالك عدة طرائق يمكن من خلالها الحصول على تقديرات جيدة لمعلمة او عدة معلمات في المجتمع الإحصائي ومن هذه الطرائق:

1- طريقة الامكان الأعظم (Method of maximum likelihood)

2- طريقة اقل تباين (Method of minimum variance)

3- طريقة المربعات الصغرى (Method of Last squares)

4- طريقة العزوم (Method of Moments)

5- طريقة اقل مربع كاي (Method of minimum χ^2) (10)

ومن اهم هذه الطرائق هي طريقة الامكان الأعظم (MLE) التي تم تناولها في بحثنا هذا.

دالة الامكان الأعظم (Maximum likelihood function)

قـدم فيشـر (Fisher) عـام 1922 طريقـة تقـدير الامكـان الاعظـم (Maximum likelihood function) (MLE) ، وكانت احدى الطرائق الاكثر اهمية في القرن العشرين لعلم الإحصاء لأنها أعطت أفضل تقدير للمعلمة من بـين عدة تقديرات ممكنة وحققت عدة معايير للمقدر منها الكفاءة (efficiency) والكفاية (sufficiency).

ان صياغة دالة الامكان الأعظم كانت عن طريق مجموعة من الأفكـار التـي اشتملت على المعلمة والاتساق والكفاءة فضلا عن التقدير (2) .

وتستخدم دالة الامكان (likelihood function) في تقدير المعلمات وهي دالة احتمال مشترك لـ K من المتغيرات العشوائية:

$$L = L(\theta_1, \theta_2, \ldots, \theta_k / data) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

ان لهذه الدالة دوراً مهماً في التقدير، اذ يمكن من خلالها ان تتطـور وتـزداد معلوماتنا حول المعلمة θ وتعتبر قناة للمعلومات التي تزودنا بها البيانات.

ولإيجاد (MLE) نستخدم المشتقة الاولى لدالة لوغاريتم الامكان مـع الاخـذ بنظر الاعتبار كل معلمة في النموذج ، ولاكثر النماذج يوجد اكثـر مـن معلمـة واحدة، وبصورة عامة نستطيع ان نكتب دالة اللوغاريتم كالأتي:

$$L = (L(\theta_1, \theta_2, \ldots, \theta_k / data)) = Log_c(L) \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

وان (k) من معادلات المشتقة للوغاريتم الامكان تكون بالشكل الاتي :

$$\left.\begin{aligned} \frac{\partial Log_e(L)}{\partial \theta_1} &= 0 \\ \frac{\partial Log_e(L)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial Log_e(L)}{\partial \theta_k} &= 0 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

وبحل المعادلات (3) نحصل على مقدرات دالة الامكان الأعظم (MLEs):

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k \qquad (11)$$

خواص مقدرات دالة الامكان الاعظم (The properties of MLEs)

هنالك عدة خواص للمقدرات التي حصلنا عليها من خلال طريقة (MLE) ومن هذه الخواص:

1- ان مقدرات دالة الامكان الاعظم هي تقديرات متسقة.

2- ان التوزيع الاحتمالي لتقدير دالة الامكان الاعظم يقتـرب مـن التوزيـع الطبيعي عندما يكون حجم العينة كبيراً.

3- إن تقدير الامكان الأعظم هو تقدير أكثر كفاءة من بين جملة تقديرات أخرى متاحة.

4- إن تقدير الامكان الأعظم ليس بالضرورة أن يكون تقديراً غير متحيز للمعلمة. (1)

الخوارزمية الجينية (Genetic Algorithm):

وهي احد أساليب الذكاء الاصطناعي الحديثة التي تعتمد على الية الانتقاء الطبيعي ونظام الجينات الطبيعية وتتضمن الخوارزمية الجينية (GA) تكوين مجتمع ابتدائي من مجموعة من الافراد اذ يمثل كل فرد بسلسة ثابتة تسمى الكرموسوم (Chromosome) وكل كروموسوم يتكون من عدد من الجينات (Genes)، ان المعلمة في النماذج الإحصائية تقابل جيناً في كرموسوم الخوارزمية وعندما تكون الجينات قيماً منفصلة مثل الاعداد الصحيحة نستطيع ايجاد حلول مثلى من خلال (GA) وعندما تكون القيم مستمرة فيمكننا ايجاد نتائج قريبة من القيمة الحقيقة. (9)

بعد تحديد حجم المجتمع وطول الكروموسوم يتم توليد مجموعة من الارقام العشوائية لتمثيل القيم الابتدائية لجينات الافراد في المجتمع كافة، والعامل الاساسي لنجاح الخورازمية الجينية (GA) هو تشفير قيم الكروموسوم ، اي ايجاد تمثيل مناسب لحل المسألة. (6)

ولاجل تكوين كروموسومات تعطي حلولاً جديدة للجيل اللاحق لم يسبق تكوينها في الاجيال السابقة وتؤدي الى الحصول على حل اكثر اقتراباً من الحل الامثل يجب ان تجري عمليتان مهمتان في الخورازمية الجينية (GA) وهما عملية التداخل الابدالي (Crossover) وعملية الطفرة (Mutation) وتجرى هاتان العمليتان بين الكروموسومات وبين جينات الكرموسوم الواحد.

تحسب كل من دالة الهدف (Objective function) وقيمة الجودة (Fitness value) وتحدد القيمة الأخيرة مقدار جودة الكروموسوم قياساً بباقي كروموسومات الجيل، ثم نجد احتمالية مساهمة كل كروموسوم في تكوين الجيل اللاحق ولاجل

انتاج جيل جديد تجرى عملية الانتقاء (Selection) لاختيار أفضل أفراد المجتمع لغرض انتاج هذا الجيل .

يستمر تكوين الاجيال الى أن يحصل على الحل الامثل المطلوب أو القريب منه وعندها نتوقف عن التوليد والاعتماد على الحل الناتج الأخير (7).

خطوات الخوارزمية الجينية المقترحة لإيجاد قيمة المقدر الذي يعظم دالة الإمكان للتوزيع الطبيعي

(Steps of the Proposed GA for finding the value of the estimator which maximizes the likelihood function for normal distribution)

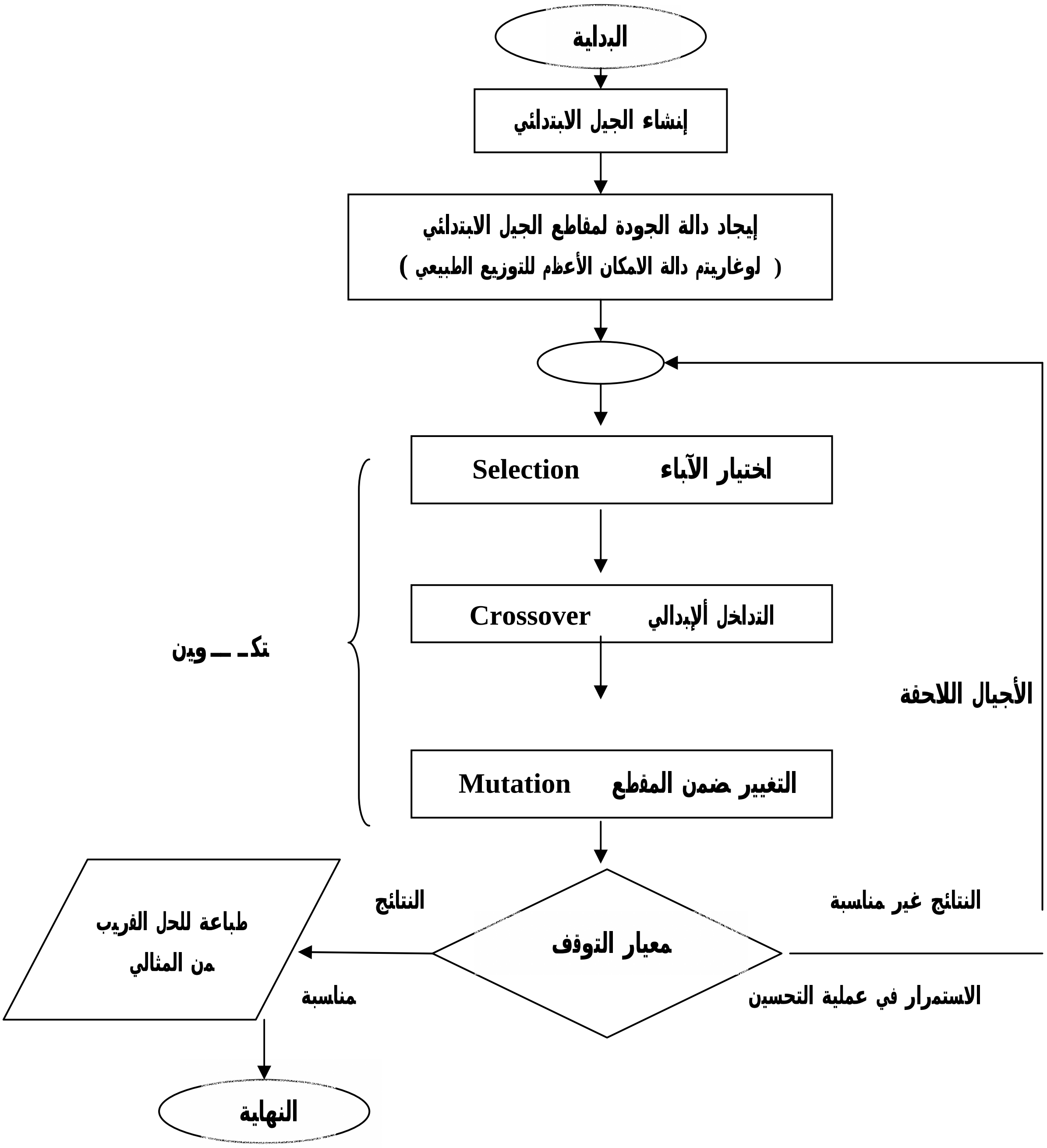
ان طريقة تعظيم دالة الامكان هي إحدى طرائق تقدير المعلمات وفي هذه الطريقة نجد قيمة المعلمة التي تحقق أعظم قيمة للدالة وقد تكون قيمة واحدة أو قيمتين ، وفي هذا البحث قمنا بتعظيم دالة الامكان للتوزيع الطبيعي عن طريق المعلمتين μ , σx أي (الوسط الحسابي ، الانحراف المعياري) اذ استخدمنا دالة الامكان للتوزيع الطبيعي في الخطوات المقترحة للخوارزمية الجينية (GA) ، والخطوات موضحة في أدناه :

1. البيانات الأولية (Initial Data): وهي قراءة لمجموعة من المتغيرات التي استخدمت في الخوارزمية :

- مصفوفة المشاهدات (T) : تمثل عناصرها مجموعة من الأرقام (المشاهدات)
- L : تمثل دالة الامكان للتوزيع الطبيعي
- S : تمثل الانحراف المعياري للتوزيع الطبيعي
- M : تمثل الوسط الحسابي للتوزيع الطبيعي
- N : عدد المشاهدات

2. إنشاء الجيل الابتدائي (Initial Generation) : في هذه الخوارزمية لدينا كروموسومان احدهما يمثل الوسط الحسابي (M) والآخر يمثل الانحراف

المعياري (S) ولقد تم وضـع القـيم الحقيقيـة (المشـاهدات) فـي جينـات الكروموسومين أي ان التشفير كان تشفيراً حقيقياً للكروموسومات .



الشكل (1-1) يبين خطوات الخوارزمية الجينية لمقدر دالة الامكان الأعظم للتوزيع الطبيعي

3. قيمة الجودة (Fitness Value) : ان قيمة الجودة في هذه الخوارزميـة تمثـل لوغاريتم دالة الامكان الأعظم للتوزيع الطبيعي ، ولقد تم استخدام اللوغـارتيم لانه الاساس في استخراج المقدارات التي تجعل الدالة اعظم ما يمكن ، وكذلك للتغلب على قيمة (exp) العالية في معادلة (likelihood) للتوزيع الطبيعي.

بعد أن وجدنا قيمة الجودة نقوم باستخدامها لإيجاد دالة الهدف (أعظم قيمـة لدالة الامكان) وذلك عن طريق إدخال هذه القيمة فـي (Toolbox) للخوارزميـة الجينية (GA) في برنامج (MATLAB7) وبعد ذلـك نحـدد عـدد المتغيـرات (المعلمات) التي ستحقق القيمة العظمى للدالة كما نختار نوع كل مـن (التـداخل ألإبدالي، الطفرة ، الانتقاء) وكذلك عدد مرات توليد الأجيال وملاحظة النتائج بعد كل توليد اذ تم توليد مابين 10 الى 100 جيل واعتمدنا على التحديد المسبق لعـدد الأجيال لبيان مدى التقرب من الحل الامثل ولقد حققت هذه الخوارزمية أعظم قيمة لدالة الامكان للتوزيع الطبيعي .

وفيما ياتي توضيح لبعض أنواع التداخل الابدالي والطفرة والانتقاء التي تم استخدامها في الخوارزمية:

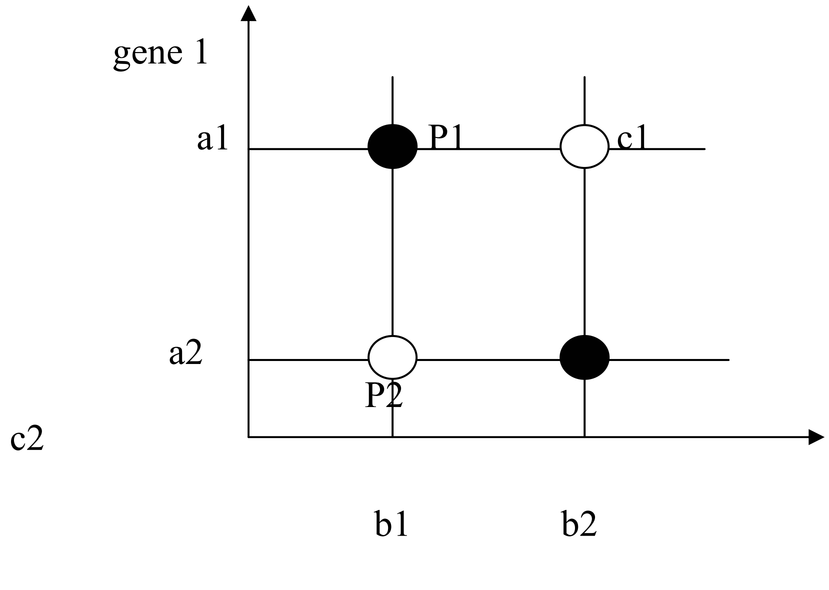
- **التداخل الابدالي المشتت (Scattered Crossover)** :

في هذا النوع من التداخل تكون الكروموسومات على شكل متجهات وباستخدام (Mask) يتم إنتاج الجيل الجديد :

P1 = [a1 , b1] , P2 = [a2 , b2] , Mask = [1 0]

c1 = [a1 , b2] , c2 = [a2 , b1]

والشكل (2-1) يوضح هذا النوع من التداخل :(5)



الشكل (2-1) التداخل الابدالي المشتت

- **التداخل الابدالي ذو نقطة القطع الواحدة (Single Point Crossover):**

في هذه الطريقة يتم اختيار عشوائي لنقطة واحدة ضمن مقطعـي الأبـوين وبعدها يتم إبدال الجينات المتقابلة لإنتاج مقطعي الأبناء والشـكل (3-1) يوضـح التداخل : (4)

Parent 1: 11001|**010** Offspring1: 11001|**111**
Parent 2: 00100|**111** Offspring2: 00100|**010**

الشكل (3-1) التداخل الابدالي ذو نقطة قطع واحدة

- **التداخل الابدالي المتوسط (Intermediate Crossover) :**

يستخدم في هذا النوع من التداخل المعادلات الاتية :

$$c1 = a * P1 + (1-a) * P2$$
$$c2 = (1-a) * P1 + a * P2$$
$$a = (1+2 * b) * r - b$$

إذ أن : b : اكبر وتساوي صفراً ، r : عدد عشوائي تتراوح قيمته بين الصـفر والواحد .

- **التداخل بالاعتماد على الحدس (Heuristic Crossover) :**

وهو عملية التداخل بين اثنين من الآباء لإنتاج الأبناء وفق المعادلات الاتية :

Offspring 1 = Best Parent + r * (Best Parent - Worst Parent)
Offspring 2 = Best Parent

إذ أن :

r : عدد عشوائي تتراوح قيمته بين الصفر والواحد . (5)

- **الطفرة الطبيعية (Gaussian Mutation) :**

تتم إضافة قيمة عشوائية من التوزيع الطبيعي على بعض الجينات المختارة من كروموسومات الآباء لإنتاج الأبناء ، ويتم تطبيق هذه الطفرة على جينـات ذات أعداد صحيحة.

- **الطفرة المنتظمة (Uniform Mutation) :**

في هذه الطفرة يتم إبدال قيمة جين مختار من الكروموسوم بقيمة عشوائية تكون محصورة بين أعلى قيمة واقل قيمة في الكروموسوم .(4)

- **الانتقاء المنتظم (Uniform Selection) :**

يتم اختيار الآباء بشكل عشوائي من التوزيع المنتظم بالاعتماد على التوقعات وعدد الآباء، وهذه الطريقة هي طريقة بحث غير مباشرة .

- **انتقاء الأشكال المنتظمة (Stochastic Uniform Selection) :**

لكل أب في هذا النوع من الانتقاء متجه ينطبق مع المتجه المناسب لاحتماله والخوارزمية تسير على طول المتجه بخطوات متساوية (خطوة واحدة لكل أب من الآباء) ، والخطوة الأولى تمثل رقما عشوائيا منتظماً قيمته اقل من طول الخطوة .

- **انتقاء عجلة الروليت (Roulette Wheel Selection):**

تعتمد هذه الطريقة على اختيار أفضل ما في المجتمع من أفراد، وذلك بعد حساب مجموع الجودة للمجتمع وحساب نسبة جودة كل فرد إلى المجموع الكلي، وتستخدم القيم الناتجة كاحتمالية لاختيار الأفراد في الأجيال اللاحقة.**(5)**

الجانب التطبيقي (Application Part):

تم تنفيذ البرنامج الذي صمم بلغة MATLAB على عدد من المشاهدات :
T=10, 20, 30, 40, 50

والجدول الاتي يوضح النتائج :

	تسلسل التوليد الذي حقق أعلى قيمة للدالة	نوع التداخل الإبدالي	نوع الطفرة	نوع الانتقاء	أعظم قيمة للدالة M.L.E.
1.	100	Scattered	Gaussian	Uniform	30.344
2.	100	Single point	=	=	23.645
3.	49	=	Uniform	=	3047.965
4.	59	=	=	=	2835.348
5.	50	=	=	=	2931.605
6.	47	=	=	=	2659.84
7.	10	Scattered	Gaussian	=	103.08
8.	2	=	=	=	263.339
9.	6	=	=	=	99.236
10	100	=	=	=	27.004
11	52	Intermediate	Uniform	=	4289.132
12	100	Scattered	Gaussian	Stochastic uniform	21.926
13	100	=	=	Roulette	22.788
14	100	Heuristic	=	=	20.34
15	100	Intermediate	Uniform	=	2707.848
16	73	=	=	=	2775.006
17	84	=	=	=	2755.836
18	99	=	=	=	2933.737
19	77	=	=	=	3512.39
20	5	scattered	Gaussian	Uniform	137.5119

نلاحـظ ان أعلـى قيمـة لدالـة الامكـان الأعظـم (M.L.E) تسـاوي (4289.132) والتي ظهرت من الجيل (52) في التوليد الثالث للخوارزمية الجينية المقترحة ، حيث كان التداخل الإبدالي من نـوع (Intermediate) أمـا الطفـرة والانتقاء فكانا من نوع (Uniform) .

وبهذا حققت الخوارزمية الجينية المقترحة تعظيم دالـة الامكـان للتوزيـع الطبيعي .

References : المصادر

1- هرمز ، أمير حنا (1990) : "الإحصاء الرياضي" ، مديرية دار الكتب والنشر، جامعة الموصل ، وزارة التعليم العالي والبحث العلمي .

2- Bennett, J. H., ed. (1990). "Statistical Inference and Analysis", Selected Correspondence of R. A. Fisher. Oxford. Univ. Press.

3- Devore, J., (2000): "Probability and Statistics", by Brooks, Cole.

4- "Gatss Information", the web Site at www.acc.umu.se/~top/travel-information.html. (2003)

5- "Genetic Algorithms" the web Site at www.cs.felk.cvut.cz/~xobitko/ga/main.html. (1998).

6- Gen, M. (2000): "Genetic Algorithms and Engineering Optimization, John Wiley and Sons, Inc.

7- Koza, J. R., (1992): "Genetic Programming", MIT Press, Cambridge, MA.

8- Levin, R. I. and Rubin, D. S., (1998): "Statistics for management", Prentice-Hell, Inc.

9- Mitchell, M. (1998): "An Introduction to Genetic Algorithms", MIT Press London.

10- Silvey, S. D., (1970): "Statistical Inference", Lowe and Brydone (Printers) Ltd, London.

11- Wonnacott, R. J. and T. H., (1985): "Introductory Statistics", by John Wiley and Sons, Inc.