تقدير معلمات توزيع رايلي المختلط باستعمال طريقة الامكان الاعظم

الباحث:نور هاشم محمد الهلالي م.د. فضاء مزهر هاشم
جامعة بغداد - كلية الادارة و الاقتصاد

المستلخص
ان التوزيعات المختلطة تُستخدم على نطاق واسع لتوفير تطبيقات ملائمة لنمذجة التوزيعات المعقدة للبيانات حول الظواهر العشوائية. تدعم النماذج المختلطة المحدودة مجموعة متنوعة من التقنيات في المجالات علمية الرئيسية مثل الفيزياء و الطب الإحصاء وغيرها ، و تلعب دورها المباشر في تحليل البيانات والاستدلال في توفير معلومات وصفية للظواهر والتجارب الاحصائية التي يكون فيها استخدام النموذج الغير مختلط غير ملائم لوصف الظاهرة اذ يتم اللجوء الى النماذج المختلطة للحصول على نتائج اكثر دقة للظاهرة تحت الدراسة.في هذا البحث تم التوصل الى ايجاد توزيع جديد مستخلص من التوزيعات الشائعة باستعمال معلمة الخلط (π) حيث يمكن انموذج توزيع مكون من عدد محدد من مركبات من نفس التوزيع و بعدة معلمات مختلفة.في هذا البحث استخدمنا توزيع رايلي الاعتيادي بمعلمة واحدة وتم خلطة بعدد محدد من مركبات نفس التوزيع باستعمال معلمة جديدة و هي معلمة الخلط و التي تبين وزن كل مركب في انموذج التوزيع المختلط و تم تطبيق الفكرة على البيانات الحقيقية وتم استعمال اسلوب تقدير معلمات الانموذج باستعمال طريقة الامكان الاعظم و تم رسم الدالة و وضعت النتائج في جداول . ان اهم ما توصل الية الباحث هو امكانية الكشف الدقيق عن مكونات المزيج باستعمال المعيار البيزي للمعلومات (BIC) التي اعطى نتائج مرضية.
الكلمات المفتاحية: التوزيع المختلط , توزيع رايلي المختلط , طريقة الامكان الاعظم , نيوتن رافسون. BIC

Estimation of parameters of the mixed Rayleigh distribution using Maximum " "Likelihood method

Mixed distributions are widely used to provide convenient applications for modeling complex distributions of data about random phenomena. Limited mixed models support a variety of techniques in major scientific fields such as physics, medicine, statistics, etc., and play their direct role in data analysis and inference in providing descriptive information for statistical phenomena and experiments in which the use of the non-mixed model is inappropriate to describe the phenomenon, as models are resorted to mixed to obtain more accurate results for the phenomenon under studyIn this paper, it was reached to find a new distribution extracted from common distributions using the mixing parameter (π), where it is possible to model a distribution consisting of a specified number of compounds of the same distribution and with several different parameters. Compounds of the same distribution using a new parameter, which is the mixing parameter, which shows the weight of each compound in the mixed distribution model. The idea was applied to the real data. An estimate of the model's . The function was drawn and parameters was found using the Maximum Likelihood method the results were placed in tables. The most important finding of the researcher is the possibility of accurate detection of the components of the mixture using the Bayesian Information Criterion (BIC), which gave satisfactory results.

Key words: mixed distribution, mixed Rayleigh distribution, Maximum Likelihood method, Raphson. BIC-Newton

١- المقدمة

توزيع رايلي من التوزيعات الاحصائية المهمة التي قدمها العالم لورد رايلي (١٨٨٠) في مشكلة في مجال ال الصوتيات .وبعد هذا تم استخدامة في مختلف مجالات العلوم و التكنلوجيا بما في ذلك الاحصاء لاحتساب عمر المكائن و في مجالات الحياة و الفيزياء لنمذجة العمليات ارتفاع الموجة و اشعاعات الصوت و الضوء. و يحوي التوزيع على نماذج تحوي على معلمتين الشكل و القياس و نموذج ذو المعلمة القياس فقط حيث سيتم دراسة النموذج الثاني في هذا البحث.

لو فرضنا ان متغير عشوائي (x) يتوزع توزيع رايلي بمعلمة واحدة حيث سيمتلك الخصائص الاحصائية الاتية:

- دالة الكثافة الاحتمالية للتوزيع

$$0<x<\infty \ , \lambda >0 f(x) = \frac{x}{\lambda^2} e^{-\frac{x^2}{2\lambda^2}}$$

المتوسط :

$$m=\sqrt{\frac{pi}{2}}\lambda$$

- دالة التجميعية للتوزيع

$$F(x)=1-e^{-\frac{x^2}{\lambda^2}}$$

- مقدر الامكان الاعظم للتوزيع

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n} t^2}{n}$$

٢- توزيع رايلي المختلط

يتم الحصول على النموذج من خلال خلط عدد محدد من مركبات التوزيع من خلال استخدام معلمة الخلط(π) حيث هذه المعلمة تمتلك الخصائص التالية :($0<\pi_i<1$)و ($\sum_{i=1}^{k}\pi_i = 1$) .و بهذه الطريقة نستطيع الحصول على التوزيعات المختلطة و بمعلمات مختلفة.و يمكن بناء نموذج التوززيع المختلط باستخدام معلمة الخلط (π_j) كالاتي:

$$f(x, \propto \pi_{k,}\lambda_k) = \sum_{i=1}^{k} \pi_i f_i(x, \lambda_i) \qquad \dots (1)$$

ان عدد معلمات الخلط هي اقل من عدد المركبات بمقدار واحد وذلك لغرض معرفة مدى مساهمة كل مركب في التوزيع المختلط.

وان الدالة التجميعية للتوزيع هي

$$F(x) = \pi_1 F_1(x) + \pi_2 F_2(x) + \dots\dots\dots + \pi_k F_k(x)$$

وبالسبة لتوزيع رايلي سيكون كالتالي:

$$f(x) = \pi_1 \frac{x}{\lambda_1^2} e^{-\frac{x^2}{2\lambda_1^2}} + \pi_2 \frac{x}{\lambda_2^2} e^{-\frac{x^2}{2\lambda_2^2}} + \cdots + \pi_k \frac{x}{\lambda_i^2} e^{-\frac{x^2}{2\lambda_i^2}} \qquad \dots (2)$$

k=i-1

i=1,2,…,k+1

و دالةالتجميعية للتوزيع هي:

$$F(x) = \pi_1 H_1(x) + \pi_2 H_2(x) + \dots + \pi_k H_i(x) \qquad \dots (3)$$

حيث ان :

$$,k+1\dots \quad \& \quad i= 1,2, H_i(x) = 1 - e^{-\frac{x^2}{2\lambda_i^2}}$$

٣- معيار المعلومات البيزي (Bayesian Information Criterion)

من المعايير الشائعة لاختبار عدد المركبات الداخلة في تركيب انموذج التوزيعات المختلطة المعيار المعلومات البيزي (Bayesian Information Criterion) الذي اقترحه العالم (Schwars-1978) و يعرف بالاختصار (BIC) الذي يتضمن عدد المعلمات المتوفرة لوغاريتم دالة الامكان (log-likelihood) و عدد المشاهدات المتوفرة في مجاميع البيانات المركبة و يعطى بالصيغة التالية :

$$BIC = -2\log(L) + \theta\log(n) \qquad \dots (4)$$

حيث ان:

$\log(L)$: تمثل الدالة اللوغاريتمية للإمكان الاعظم التي يتم استعمالها لتقدير الانموذج.

θ: تمثل عدد المعلمات المقدرة

n: عدد المشاهدات

٤- التقدير:

التقدير من اعم المفاهيم في ايجاد المقدرات للمعلمات المجهولة في المجتمع بالتعامل المعلومات المتوفرة من العينة و بالتالي فان المقدر يكون دالة في المشاهدات العشوائية وهناك العديد من الطرق التي يمكن من خلالها ايجاد تقدير هذه المعلمات سنستعمل في بحثنا هذا طريقة الامكان الاعظم لايجات مقدرات الانموذج توزيع رايلي المختلط .

٤-١ طريقة الامكان الاعظم (Maximum Likelihood Estimation)

تعتبر هذه الطريقة من اهم الطرائق في عملية التقدير و اكثرها استعمالاً , الطريقة تعتمد على ايجاد مقدر يجعل دالة الامكان الاعظم في نهايتها العظمى.

و تتميز هذه الطريقة بخصائص جيدة كخاصية الكفاية و الاتساق و امتلاكها اقل تباين خطاء و تزداد كفاءة هذه الطريقة عندما يكون حجم العينة كبير اذ تكون غير متحيزة.

و فكرة هذا الاسلوب هو جعل لوغاريتم دالة الامكان (log-Likelihood function) عند نهايتها العظمى و لغرض الحصول على المقدرات يتم اجراء التفاضل الجزئي للمعلمة المراد تقديرها و مساوات الناتج بالصفر في كل حالة .

فاذا كان لدينا (x) متغير عشوائي يتوزع توزيع رايلي مختلط بعدد محدد من المركبات و حسب المعادلة رقم (2) :

$$f(x;\theta) = \pi_1 \frac{x}{\lambda_1^2} e^{-\frac{x^2}{2\lambda_1^2}} + \pi_2 \frac{x}{\lambda_2^2} e^{-\frac{x^2}{2\lambda_2^2}} + \pi_3 \frac{x}{\lambda_3^2} e^{-\frac{x^2}{2\lambda_3^2}} + \cdots + \pi_k \frac{x}{\lambda_k^2} e^{-\frac{x^2}{2\lambda_k^2}} \qquad \dots (5)$$

و للحصول على تعظيم الدالة نظرب الدالة ب (n) من المرات و كما في المعادلة التالية:

$$\prod_{i=1}^{n} f(x;\theta) = \prod_{i=1}^{n} (\pi_1 \frac{x_i}{\lambda_1^2} e^{-\frac{x_i^2}{2\lambda_1^2}} + \pi_2 \frac{x_i}{\lambda_2^2} e^{-\frac{x_i^2}{2\lambda_2^2}} + \pi_3 \frac{x_i}{\lambda_3^2} e^{-\frac{x_i^2}{2\lambda_3^2}} + \cdots + \pi_k \frac{x_i}{\lambda_k^2} e^{-\frac{x_i^2}{2\lambda_k^2}}) \quad . \quad . \quad . \quad (6)$$

وبذلك ستكون دالة الامكان بالشكل التالي :

$$L(\theta|x) = \prod_{i=1}^{n}\sum_{j=1}^{k} \pi_j \{f_j(x_i;\lambda_j)\} \quad . \quad . \quad . \quad (7)$$

و لصعوبة حل المعادلات و الوصول الى المقدرات سوف يتم اعتماد اسلوب خوارزمية نيوتن رافسون العددية باستعمال الحاسوب.

٥- **متعدد متغيرات يوتن رافسون (Multivariate Newton-Raphson Method)**

يعد اسلوب نيوتن رافسون في حل المعادلات الغير خطية اسلوب تكراري و لحل تلك المعادلات لابد من تحقق المعادلة التالية :

$$\begin{pmatrix} f_1(\theta_1,\dots,\theta_k) \\ f_2(\theta_1,\dots,\theta_k) \\ \vdots \\ f_k(\theta_1,\dots,\theta_k) \end{pmatrix} = f_\theta = 0$$

فاذا كان (θ) تمثل فضاء المعلمات اي :

$$\theta = (\pi_1, \pi_2, \dots, \pi_{k-1}, \lambda_1, \lambda_2, \dots, \lambda_k)$$

اذا في البدء علينا ان نجد تخمينا للمعلمة الجدية (θ_{r+1}) باستعمال (Taylor-expanding) لدالة $(f(\theta_r))$ و كما يلي:

$$f(\theta_{r+1} \approx f(\theta_r) + J(\theta)_r = 0) \quad . \quad . \quad . (2.76)$$

حيث ان(J) هو مصفوفة من الدرجة (2k-1*2k-1) حيث تكون قيمتها مساوية الى :

$$J_{ij} = \partial f_i / \partial \theta_j \quad . \quad . \quad . (8)$$

اي ان :

$$J_{n*n} = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \cdots \\ \frac{\partial f_2}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad . \quad . \quad . \quad (10)$$

إذا توظيف خوارزمية نيوتن للإيجاد التقديرات الجديدة و حسب الصيغة ادناه

$$\theta_{r+1} = \theta_r - \frac{f(\theta_r)}{\acute{f}(\theta_r)} \quad . \quad . \quad . \quad (11)$$

٦- **التطبيق العملي**

تم في هذا البحث الاعتماد على البيانات الماخوذة من جهاز الارسال للبيانات المستعمل في تقنيات الاتصالات واحتساب اوقات توقف لثلاث اجهزة من نفي النوع في ثلاث اماكن مختلفة و اظهرت النتائج ان البيانات تعود الى ثاث مركبات حسب اختبار (BIC) وكما فيالجدول التالية .

Test for the number of components by BIC

جدول رقم (١) يبين اختبار الكشف عن المركبات الداخلة في تكوين الانموذج				
5coms	4coms	3coms	2 coms	1 coms
٥٠٠٥,٨٥٥	٤٨٤٨,٧٨٥	٤٨٤١,٩١٨	٤٨٥١,٠١٦	٤٨٦٣,٥٨٣

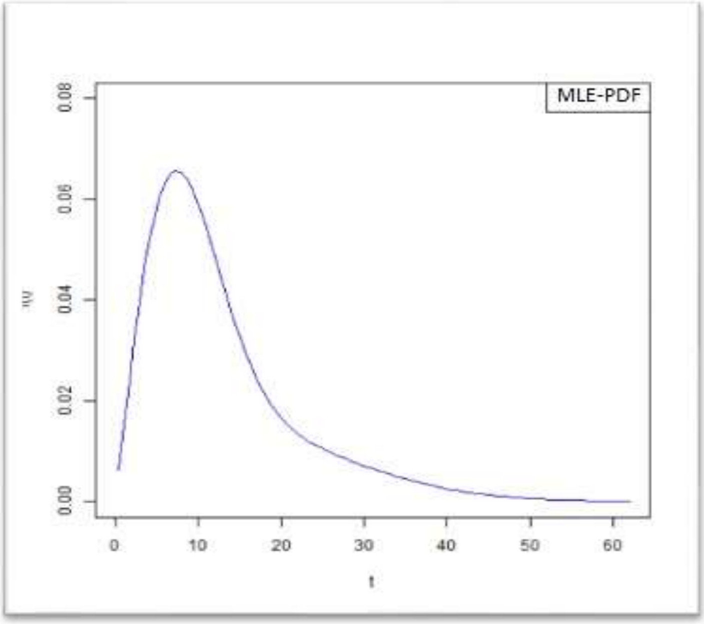
٧- **تحليل جدول الاختبار**

تبين من خلال الاختبار ان في حقل ثلاث مركبات (3coms) تظهر اقل قيمة لقيم الاختبار وهذه الدلالة على ان البيانات تتبع ثلاث ركبات و باستخدام الحاسوب.

بعد التوصل الى عدد مركبات المكونة لانموذج المختلط من توزيع رايلي يمكن الوصل الى تقدير معلمات انموذج ثلاث مركبات باستعمال طريقة الامكان الاعظم و ثبتت النتائج في الجدول التالي:

جدول رقم (2) تقدير معلمات انموذج ثلاث مركبات من توزيع رايلي المختلط							
	n	λ_1	λ_2	λ_3	π_1	π_2	π_3
MLE		16.34	7.33	6.63	0.3386579	0.3332974	0.3280447

٨- **تحليل النتائج**
الجدول اعلاه يبين نتايج التي تم الحصول عليها من خلال استخدام الحاسوب و ادخال معادلة التقدير لتوزيع رايلي المختلط بعد ادخال جميع قيم البيانات التي تم الحصول عليها من مصادر حقيقة و اضهرت النتائج يمكن الاعتماد على طريقة الامكان الاعظم لتقدير معلمات التوزيعات المختلطة .



شكل رقم (١) يبين سلوك دالة (pdf) لانموذج ثلاث مركبات من توزيع رايلي المختلط

٩- **الاستنتاجات** :
يمكن تلخيص اهم الاستنتاجات التي تم الحصول عليها في اربع نقاط رئيسية:
١- امكانية استعمال معيار المعلومات البيزي في الكشف على عدد المركبات المكون للمزيج من المشاهدات .
٢- امكانية الحصول على تقدير وزن كل مركب من مركبات المكون للانموذج التوزيع المختلط .
٣- نلاحظ وزن المركب الاول هو اعلى وزن من باقي المركبات و ان المركب الثالث هواقل وزناً.
٤- امكانية قبول النتائج اي انها حقيقة.

1. Abraham, D. A., Gelb, J. M., & Oldag, A. W. (2010, May). K-Rayleigh mixture model for sparse active sonar clutter. In *OCEANS'10 IEEE SYDNEY* (pp. 1-6). IEEE.
2. Afify, W. M. (2011). CLASSICAL ESTIMATION OF MIXED RAYLEIGH DISTRIBUTION IN TYPE IProgressive CENSORED. *Journal of Statistical Theory and Applications ISSN*, *1538*, 7887.
3. Arcidiacono, P., & Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the EM algorithm. *Econometrica*, *71*(3), 933-946.
4. Arcidiacono, P., & Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the EM algorithm. Econometrica, 71(3), 933-946.
5. Ateya, S. F. (2014). Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. *Statistical papers*, *55*(2), 311-325.
6. Davenport, J. W., Bezdek, J. C., & Hathaway, R. J. (1988). Parameter estimation for finite mixture distributions. *Computers & mathematics with applications*, *15*(10), 819-828.

7. EA, E. (2007). Estimation of parameters of mixed generalized exponentionally distributions from censored type I samples. Journal of applied sciences Research, 3(12), 1696-1700.
8. Farrell, P. J., Ehsanes Saleh, A. K., & Zhang, Z. (2011). Methods of moments estimation in finite mixtures. Sankhya A, 73(2), 218-230.
9. Farrell, P. J., Saleh, A. M. E., & Zhang, Z. (2011). Methods of moments estimation in finite mixtures. *Sankhya A*, *73*(2), 218-230.
10. Feroze, N., & Aslam, M. (2019). Three-Component Mixture of Rayleigh Model Under Doubly Censored Samples: A Bayesian Look. Communications in Mathematics and Statistics, 7(4), 417-443.