بناء التوزيع المختلط ايلاموجيا – رايلي لتقدير دالة البقاء

الاستاذ المساعد الدكتورريسان عبدالامام زعلان
المدرس المساعد منتظرجمعة مهدي
جامعة البصرة / كلية الإدارة والاقتصاد /قسم الإحصاء

تواريخ البحث
تاريخ تقديم البحث : 2024/10/15
تاريخ قبول البحث :2024/12/1
تاريخ رفع البحث على الموقع:
2024/9/15

المستخلص :

تم في هذا البحث بناء توزيع احتمالي جديد من خلال خلط توزيعين من توزيعات الفشل وهما توزيع ايلاموجيا وتوزيع رايلي ليكون التوزيع الجديد (ايلاموجيا – رايلي) يتألف من ثلاث معالم من ضمنها معلمة الخلط. وتم استعمال التوزيع المختلط الجديد في تقدير دالة البقاء لمرضى سرطان الثدي اذ تعتبر التوزيعات المختلطة اكثر مرونة وكفاءة من التوزيعات المفردة في تمثيل البيانات المعقدة ومنها بيانات البقاء لذا تمت مقارنة التوزيع المختلط الجديد مع التوزيعات المفردة رايلي و ايلاموجيا واثبت التوزيع المختلط كفاءته اذ كان افضل في تمثيل بيانات البقاء من التوزيعات المفردة حسب معيار متوسط مربعات الخطأ MSE .

الكلمات المفتاحية: توزيع رايلي، توزيع ايلاموجيا، التوزيع المختلط، مقدرات الإمكان الاعظم، معيار معلومات اكايكي AIC

Build the Amixture Distribution Ailamuji-Raleigh for Estimating Survival Function

Assistant professor Dr. Risan Eabdalamam Zaelan

Assistant lecturer Muntazar Jameuh Mahdi

College of Admin & Economic / Statistics Dep./ University of Basrah

Abstract :

In this research, two distributions that are considered failure distributions were mixed, namely the Elamogya distribution and the Rayleigh distribution, as the new distribution became composed of four parameters, including the mixing parameter, and the new mixed distribution was used to estimate the survival function for breast cancer patients, as the mixed distributions are considered more flexible and efficient than the single distributions in representing complex data, including survival data. Therefore, the new mixed distribution was compared with the single distributions, Rayleigh and Elamogya, and the mixed distribution proved its efficiency if it was better in representing survival data than the single distributions according to the mean square error (MSE) criterion.

Keywords: Rayleigh distribution, Elamogya distribution, Mixed distribution, Maximum likelihood estimators, Akaike information criterion (AIC)

المقدمة :

من اجل زيادة الدقة وتقليل الخطا في التحليل الاحصائي فان من الضروري ان تكون البيانات المستعملة للظواهر المختلفة في التحليل اكثر ملائمة للتوزيع المستعمل في الدراسة لذا يتم استعمال توزيعات مفردة اساسية ومعروفة لتمثيل هذه الظواهر ولكن بعض الظواهر تأخذ اشكال ومنحنيات مختلفة قد لا تتلائم مع هذه التوزيعات وذلك لان بيانات هذه الظواهر عبارة عن مزيج من مجتمعين او اكثر لذا كان لابد من البحث عن توزيعات تكون اكثر كفاءة ومرونة لتمثيل هذا النوع من البيانات حتى تكون عملية التحليل اكثر دقة وكفاءة. ومن اجل الوصول الى ذلك قام الباحثون بخلط توزيعين او اكثر من اجل الحصول على توزيع جديد بخصائص إحصائية جديدة من اجل تمثيل هذه البيانات المعقدة بحيث تكون افضل مقارنة بالتوزيعات المفردة عند اجراء تحليل المطابقة او استعمالها في التطبيق على البيانات المعقدة.

مشكلة الدراسة

ان التمثيل الأمثل للتوزيعات الاحتمالية للبيانات الاحصائية كانت وما زالت تستقطب العديد من الباحثين وفي مختلف المجالات ونتيجة لوجود بيانات إحصائية تكون اكثر تعقيد من غيرها من البيانات الأخرى بحيث تكون التوزيعات الكلاسيكية المعروفة لا تمثلها بصورة امثل مما حذى بالباحثين بالبحث عن توزيعات تكون اكثر ملائمة لهذه البيانات لذا لجأ الباحثون الى خلط توزيعين او اكثر بحيث يتم الحصول على توزيع مختلط ذي خصائص إحصائية تختلف عن التوزيعات المفردة التي تم استعمالها في خلط التوزيع الجديد كذلك تكون هذه التوزيعات مرنة وقادرة على تمثيل البيانات المعقدة بصور اكثر مثالية من التوزيعات المفردة لذا تم في هذا البحث خلط توزيع رايلي مع توزيع ايلاموجيا من اجل تقدير دالة البقاء لمرضى سرطان الثدي.

هدف الدراسة :

بناء توزيع احتمالي مختلط جديد هو توزيع (ايلاموجيا –رايلي) وإيجاد خصائصه الإحصائية ومقارنته مع التوزيعات المفردة رايلي وايلاموجيا في تقدير دالة البقاء لمرضى سرطان الثدي.

توزيع ايلاموجيا : Ailamujia Distribution [9],[6]

يعتبر هذا التوزيع احد اهم التوزيعات الاحتمالية المستمرة والذي تم تقديمه من قبل الباحث Lvet at عام 2002 حيث يستخدم هذا التوزيع لتحليل انواع مختلفة من البيانات منها الهندسية و الطبية وتكتب دالة الكثافة الاحتمالية للتوزيع (pdf) بالصيغة التالية :

$$f(x) = 4\theta^2 x e^{-2x\theta} \quad , X > 0 \ , \theta > 0$$

اذ ان θ هي معلمة القياس.

اما دالة التوزيع التراكمي(CDF) فتكتب بالشكل التالي :

$$F(x) = 1 - e^{-2\theta x}(1 + 2\theta x)$$

اما دالة البقاء للتوزيع والتي تمثل احتمال بقاء الفرد على قيد الحياة بعد مرور فترة زمنية محددة ويرمز لها بالرمز $S(t)$ ويمكن صياغتها بالشكل التالي:

$$S(x,\theta) = 1 - F(x) = 1 - [1 + e^{-2\theta x}(1 + 2\theta x)] = e^{-2\theta x}(1 + 2\theta x)$$

اما دالة المخاطرة فتكتب بالشكل التالي :

$$h(x,\theta) = \frac{f(x)}{s(x)} = \frac{\left(1 + 4\theta^2 x e^{-2x\theta} 2\theta x\right)}{e^{-2\theta x}(1 + 2\theta x)}$$

وان كل من توقع وتباين التوزيع يكتبان كالاتي على التوالي:

$$\mu_x = E(x) = \frac{1}{\theta}$$
$$\sigma_x{}^2 = v(x) = \frac{1}{2\theta}$$

توزيع رايلي : Rayleigh distribution[7],[3]

تم تقديم هذا التوزيع من قبل اللورد رايلي عام 1880 حيث يلعب هذا التوزيع دور مهم في تحليل بيانات البقاء وتحليل المعولية وغيرها من البيانات ويعتبر من التوزيعات الاحتمالية المستمرة الاساسية والمهمة وتكتب دالة الكثافة الاحتمالية للتوزيع (pdf) بالصيغة التالية:

$$f(x) = \frac{e^{-\frac{x^2}{2\sigma^2}} x}{\sigma^2} \quad , \quad x > 0 \quad , \quad \sigma >$$

حيث ان : σ هي معلمة القياس

وان دالة التوزيع التراكمي كما يلي :

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

وان دالة البقاء تعرف كما يلي :

$$S(x) = 1 - F(x) = 1 - [1 - e^{-\frac{x^2}{2\sigma^2}}] = e^{-\frac{x^2}{2\sigma^2}}$$

وان دالة المخاطرة تأخذ الصيغة التالية :

$$h(x) = \frac{f(x)}{s(x)} = \frac{\frac{e^{-\frac{x^2}{2\sigma^2}} x}{\sigma^2}}{e^{-\frac{x^2}{2\sigma^2}}} = \frac{x}{\sigma^2}$$

خلط التوزيعات الاحتمالية : **Mix probability distributions** [5],[4]

يطلق على التوزيع الخليط (**mixture distribution**) وهو عبارة عن تجميع لمكونات غير متجانسة لتوزيعات احتمالية احصائية ، وان هذا الخلط يكون عندما يتم سحب عينة من مجتمعات تكون غير متجانسة وبدوال احتمالية مختلفة او متشابهة وتتألف من معلمات مختلفة ولكل مجتمع جزئي . ويتم خلط التوزيعات وفق الصيغة التالية :

$$g(x) = \sum_{i=1}^{k} p_i \, fi(x)$$

اذ ان $\mathbf{p_i}$ تمثل نسبة المجتمع الجزئي **i**. اذ ان $\sum \boldsymbol{p_i} = \mathbf{1}$

وان **fi (x)** الدوال الاحتمالية المختلفة اذ ان **i=1,2,..,k** حيث ان **k** هي عدد الدوال الاحتمالية.

التوزيع المختلط (ايلاموجيا- رايلي) :

لذلك، يمكن كتابة دالة **pdf** لنموذج الخليط بالشكل التالي:

$$f(x) = p4\theta^2 x e^{-2x\theta} + (1-p)\frac{e^{-\frac{x^2}{2\sigma^2}}x}{\sigma^2}$$

في حين يمكن كتابة كل من الدالة التوزيع التراكمية ودالة البقاء ودالة المخاطرة على التوالي كما يلي:

دالة التوزيع التراكمية :

$$F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}(1-p) + e^{-2x\theta}p(1+2x\theta)$$

دالة البقاء :

$$S(x) = e^{-\frac{x^2}{2\sigma^2}}(1-p) + e^{-2x\theta}p(1+2x\theta)$$

دالة المخاطرة :

$$h(x) = \frac{x(-e^{2x\theta}(-1+p) + 4e^{\frac{x^2}{2\sigma^2}}p\theta^2\sigma^2)}{(-e^{2x\theta}(-1+p) + e^{\frac{x^2}{2\sigma^2}}p(1+2x\theta))\sigma^2}$$

وكذلك يتم ايجاد بعض خصائص التوزيع الاساسية كالاتي:

الوسط الحسابي:

$$\mu_x = E(x) = \frac{2p + \sqrt{2\pi}\theta\sigma - p\sqrt{2\pi}\theta\sigma}{2\theta}$$

التباين :

$$V(x) = \frac{3p - 2p^2 + 2(-1+p)p\sqrt{2\pi}\theta\sigma - (-1+p)(4+(-1+p)\pi)\theta^2\sigma^2}{2\theta^2}$$

الانحراف المعياري :

$$SD = \frac{\sqrt{\frac{3p - 2p^2 + 2(-1+p)p\sqrt{2\pi}\theta\sigma - (-1+p)(4+(-1+p)\pi)\theta^2\sigma^2}{\theta^2}}}{\sqrt{2}}$$

معامل الالتواء :

$$SK = \frac{-12p+18p^2-8p^3+3(-1+p)p(-3+4p)\sqrt{2\pi}\theta\sigma-12(-1+p)p(2+(-1+p)\pi)\theta^2\sigma^2+2(-1+p)\sqrt{2\pi}(-3+\pi+p(6+(-2+p)\pi))\theta^3\sigma^3}{\theta\sqrt{\frac{6p-4p^2+4(-1+p)p\sqrt{2\pi}\theta\sigma-2(-1+p)(4+(-1+p)\pi)\theta^2\sigma^2}{\theta^2}}(-3p+2p^2-2(-1+p)p\sqrt{2\pi}\theta\sigma+(-1+p)(4+(-1+p)\pi)\theta^2\sigma^2)}$$

معامل التفلطح

$$KU=\frac{\begin{array}{c}(30p-48p^2+36p^3-12p^4+ \\ 12(-1+p)p(2+p(-3+2p))\sqrt{2\pi}\theta\sigma-6(-1+p)p(3\pi+p(8+(-9+6p)\pi))\theta^2\sigma^2+ \\ 12(-1+p)p\sqrt{2\pi}(-2+\pi+p(4+(-2+p)\pi))\theta^3\sigma^3-(-1+p)(32+3(-1+p)\pi(8p+ \\ (-1+p)^2\pi))\theta^4\sigma^4)\end{array}}{\begin{array}{c}(-3p+2p^2-2(-1+p)p\sqrt{2\pi}\theta\sigma^2+ \\ (-1+p)(4+(-1+p)\pi)\theta^2\sigma^2)\end{array}}$$

معيار معلومات اكايكي Akaike Information Criterion (AIC) [2]:

ويستعمل لتقدير خطأ التنبا في النماذج الاحصائية أي انه يستعمل لتحديد جودة كل نموذج تم تطبيقه على عينة من البيانات وعن طريق هذا المعيار يمكننا اختيار النموذج الأفضل من بين هذه النماذج المطبقة على عينة البيانات وتم تسمية هذا المعيار بهذا الاسم على العالم الذي قام باشتقاقه وهو Hirotough Akaike سنة 1971.

لنفترض انه لدينا نموذج معينا وتم تطبيقه عينة معينة وعلى افتراض ان عدد المعلمات المقدرة في هذا النموذج هي p وان $L(\hat{\theta}/t)$

تمثل لوغاريتم الإمكان الأعظم لمشاهدات العينة فان الصيغة الرياضية لهذا المعيار هي :

$$\boldsymbol{AIC = -2L(\hat{\theta}/t) + 2p}$$

اذ ان النموذج الذي يملك اقل AIC هو النموذج الأفضل.

معيار معلومات اكايكي المتسق Consistent Akaike Information Criterion (AICc)[2]:

اقترح هذا المعيار من قبل Sugiura عام 1974 ويتميز بانه يميل الى ان يكون اكثر دقة من معيار معلومات اكايكي وخاصة في العينات الصغيرة وصيغته الرياضية هي :

$$AICc = -2L(\hat{\theta}/t) + \frac{2p}{2-p-1}$$

اذ تشير ال n الى حجم العينة اما ال p فتشير الى عدد المعلمات لاحظ انه كلما اقتربت قيمة n من المالانهاية تقترب قيمة AIICc الى قيمة AIC وكلما كانت قيمة n اكبر من p كانت المشكلة في معيار AIC اقل.اذ ان النموذج الذي يملك اقل AICc هو الأفضل.

معيار معلومات بيز Bayesian Information Criterion[2]:

تم تطوير هذا المعيار(BIC) من قبل Gideon E.SCHWARZ عام1978 ويستعمل للمقارنة بين النماذج اذ ان النموذج الذي يملك اقل معيار هو الافضل والصيغة الرياضية لهذا المعيار هي :

$$BIC = -2L(\hat{\theta}/t) + PLog(n)$$

اذ ان n هي حجم العينة وان ال p هي عدد المعلمات وان $L(\hat{\theta}/t)$ هي لوغاريتم دالة الإمكان.

مقدرات الامكان الاعظم Likelihood Function[8]

تعد طريقة الإمكان الأعظم الشائعة الاستعمال لتقدير معلمات دوال التوزيعات الاحتمالية، وقد تم استعمال دالة الإمكان الأعظم (Likelihood Function) لأول مرة من قبل الباحث Fisher [27] عام 1922 ويفترض هذا الاسلوب في التقدير بان المعلمة المراد تقديرها هي قيمة ثابتة (Fixed value) غير معلومة مما ميز طريقه الإمكان الأعظم امتلاكها لخاصية عدم التغاير او الثبات (Invariant) مما جعلها طريقه مفضله اضافه لانها تعطى مقدرات غالبا ما تكون غير متحيزة (Unbiased)عندما يكون حجم العينة كبيرا وامتلاكها لاقل تباين ممكن وكافية (Sufficient) ووحيدة (Unique).

ولتقدير معلمات توزيع (ايلاموجيا- رايلي) (θ, σ, p) وبفرض ان دالة الكثافة الاحتمالية هي $f(x_i|\theta, \sigma, p)$ عندها تكون دالة الإمكان (L) وفق الصيغة الاتية :

$$L = f(x_1, x_2, \dots, x_n|\theta, \sigma, p) = \prod_{i=1}^{n} f(x_i|\theta, \sigma, p)$$

$$\prod_{i=1}^{n} f(x_i|\theta, \sigma, p) = \prod_{i=1}^{n} \left(p4\theta^2 x e^{-2x\theta} + (1-p)\frac{e^{-\frac{x^2}{2\sigma^2}}x}{\sigma^2}\right)$$

ولتقدير معلمات دالة الإمكان باخذ اللوغاريتم الطبيعي لطرفي المعادلة ():

$$Log\ L = \sum_{i=1}^{n} Ln[f(x_i|\theta,\sigma,p)] \qquad ()$$

$$Log\ L = \sum_{i=1}^{n} \text{Log}\left[4e^{-2\theta x_i}p\theta^2 x_i + \frac{e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i}{\sigma^2}\right]$$

وللحصول على دالة الإمكان في نهايته العظمى يتم اشتقاق الداله بالنسبة الى المعلمات المجهولة:

$$\frac{\partial Log\ L}{\partial\theta} = \sum_{i=1}^{n} \frac{8e^{-2\theta x_i}p\theta x_i - 8e^{-2\theta x_i}p\theta^2 x_i^2}{4e^{-2\theta x_i}p\theta^2 x_i + \frac{e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i}{\sigma^2}}$$

$$\frac{\partial Log\ L}{\partial\sigma} = \sum_{i=1}^{n} \frac{-\frac{2e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i}{\sigma^3} + \frac{e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i^3}{\sigma^5}}{4e^{-2\theta x_i}p\theta^2 x_i + \frac{e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i}{\sigma^2}}$$

$$\frac{\partial Log\ L}{\partial p} = \sum_{i=1}^{n} \frac{4e^{-2\theta x_i}\theta^2 x_i - \frac{e^{-\frac{x_i^2}{2\sigma^2}}x_i}{\sigma^2}}{4e^{-2\theta x_i}p\theta^2 x_i + \frac{e^{-\frac{x_i^2}{2\sigma^2}}(1-p)x_i}{\sigma^2}}$$

وبحل المعادلات بعد مساواتها بالصفر نحصل على مقدرات المعلمات المطلوبة .

وبعد حل المعادلات بالاعتماد على الطرق العددية لايجاد المقدرات $(\hat{\theta}, \hat{\sigma}, \hat{p})$ حيث تم تطبيق طريقة نيوتن رافسون Newton's Raphson method لايجاد حلول لمنظومة المعادلات اعلاه .

يمكن الحصول على مقدر دالة البقاء باستعمال طريقة الإمكان الأعظم كما في المعادلة التالية :

$$\hat{S}(x) = e^{-\frac{x^2}{2\hat{\sigma}^2}}(1-\hat{p}) + e^{-2x\hat{\theta}}\hat{p}\left(1+2x\hat{\theta}\right)$$

الجانب العملي :

تم جمع البيانات الحقيقية من السجل الخاص بالمرضى المسجلين في مركز الاورام في محافظة صلاح الدين والذين أدت بهم الاصابة بالمرض الى الوفاة، حيث تم اخذ عينة عشوائية بحجم 95 مريضا وتحديد فترة بقائهم على قيد الحياة لحين الوفاة (بالأشهر) وادراجها في الجدول التالي:

جدول (1) مدة بقاء المريض بسرطان الثدي على قيد الحياة من تاريخ تسجيل الحالة ولحين الوفاة (بالأشهر) **[1]**

ti	ti	ti	ti	ti	ti	ti	ti	ti	ti
0.23	0.60	0.77	0.87	1.83	1.83	2.03	2.13	2.30	2.47
2.50	3.10	3.10	3.23	3.33	3.37	3.43	3.53	3.67	4.20
4.40	4.90	5.00	5.00	5.07	5.20	5.30	5.47	5.50	5.97
5.97	6.00	6.10	6.10	6.20	6.23	6.73	6.83	6.93	7.00
7.07	7.30	7.43	7.50	7.57	8.00	8.07	8.83	8.87	8.90
8.93	9.20	9.23	9.23	9.30	9.87	9.93	10.10	10.17	10.17
10.36	10.70	10.73	10.80	10.93	11.50	11.87	11.97	12.03	13.27
12.33	13.30	13.80	14.03	14.33	14.80	15.10	15.67	15.83	16.43
16.57	16.70	16.87	17.07	17.33	17.40	18.03	18.07	20.27	20.80
21.17	21.23	23.20	23.90	26.03					

المصدر(1) (السامرائي, **حاتم (2022)**

اذ تم استخراج الإحصاءات الوصفية للعينة وكما يلي :

جدول (2) يبين أبرز احصاءات العينة للبيانات الحقيقية

mean	**9.41589**
Variance	**34.9849**
skewness	**0.682298**
kurtosis	**2.80327**
median	**8.83**
StandardDeviation	**5.9148**

ثم تم استعمال طريقة الإمكان الأعظم لتقديرمعلمات التوزيعات الثلاث (ايلاموجيا , رايلي , التوزيع المختلط ايلاموجيا-رايلي) وكالتالي :

جدول (3) قيم المعلمات المقدرة بالنسبة لتوزيعات المقارنة

dist	θ	σ	p
E-R	**0.207474**	**8.80187**	**0.260883**
E	**0.106203**	_	_
R	____	**7.86269**	__

اذ ان : **E** هو مختصر ل توزيع ايلاموجيا وان **R** هو مختصر لتوزيع رايلي وان **E.-R** هو مختصر التوزيع المختلط.

لذا تم استعمال المعلمات المقدرة في حساب المعايير التالية من اجل استخراج اختبارات حسن المطابقة

جدول (4) اختبارات حسن المطابقة للبيانات الحقيقية بالنسبة لتوزيعات المقارنة

Goodness of Fit tests	Elamoga- Rayliegh		Elamoga		Rayliegh	
TEST	**Statistic**	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**
Cramér-von Mises	**0.024761**	**0.990075**	**0.055403**	**0.842882**	**0.309528**	**0.0545464**
Anderson-Darling	**0.190575**	**0.992788**	**0.387865**	**0.860338**	**2.19109**	**0.0547959**
Kolmogorov-Smirnov	**0.042337**	**0.993006**	**0.064526**	**0.799654**	**0.109533**	**0.0541936**

اذ نلاحظ بان التوزيع المختلط كان الأكثر ملائمة للبيانات من التوزيعات الأخرى وحسب المعايير الثلاثة لذا تم استعمال المعلمات للتوزيعات الثلاث في تقدير دالة البقاء للبيانات الحقيقية والمقارنة بين التوزيعات الثلاثة وحسب المعايير الثلاثة التالية :

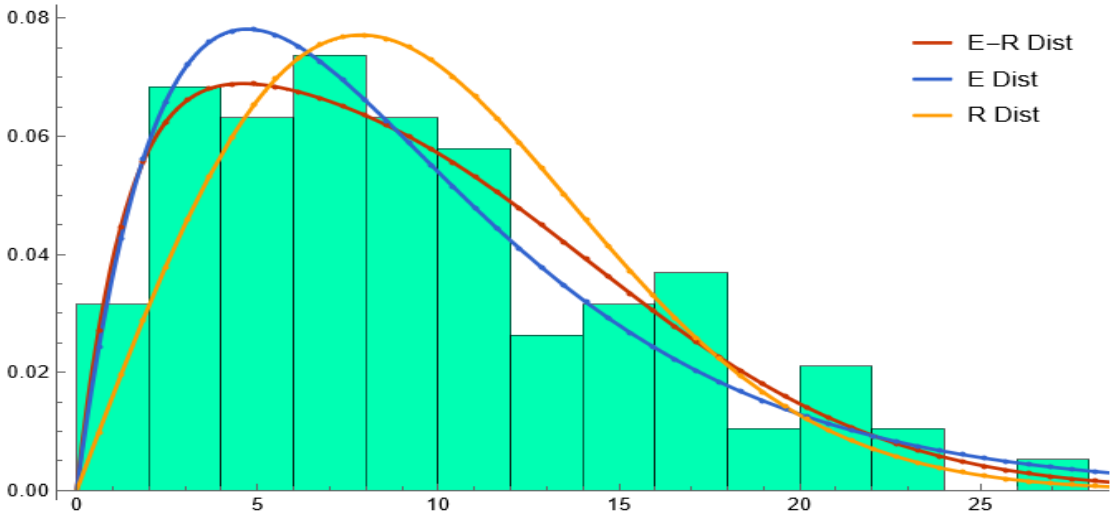
جدول (5) معايير مقارنة التوزيعات بالنسبة لتوزيعات المقارنة

Dist	AIC	AICc	BIC
E-R	**593.392**	**593.656**	**596.5**
E	**594.888**	**594.931**	**597.442**
R	**599.782**	**594.931**	**602.336**

اذ نلاحظ ان التوزيع المختلط كان الأفضل حسب المعايير الثلاث.

وكما موضح بشكل التالي :

الشكل(1) رسم دالة الكثافة الاحتمالية لتوزيعات المقارنة بالنسبة للبيانات الحقيقية



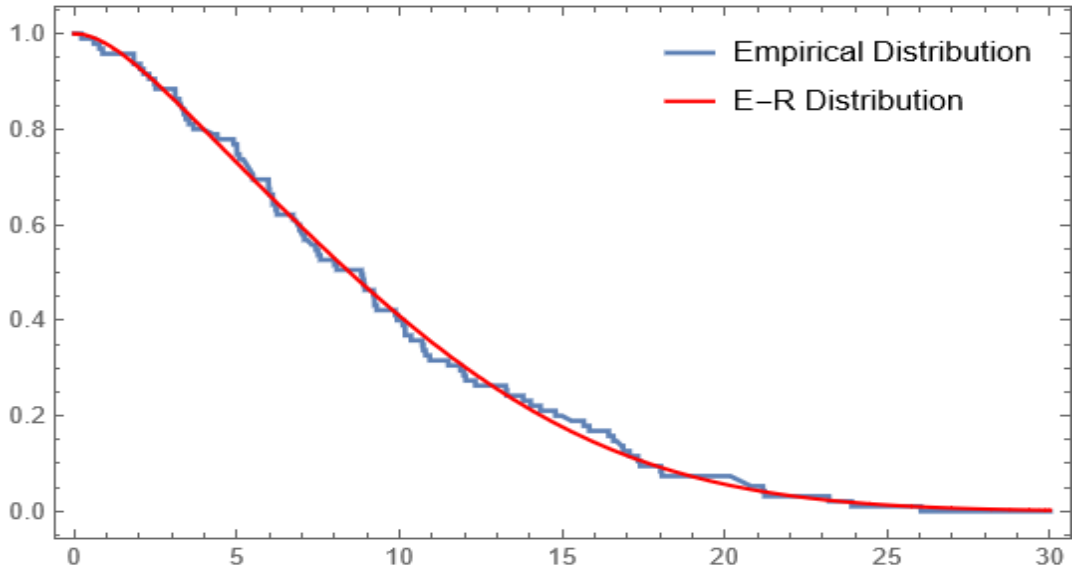
اذ نلاحظ بان التوزيع المختلط كان افضلها لذا تم استعماله في تقدير دالة البقاء لمرضى سرطان الثدي وكما يلي :

جدول (6) دالة البقاء المقدرة باستعمال توزيع (ايلاموجيا – رايلي) لبياناتمرضى سرطان الثدي

ti	SF	ti	SF	ti	SF	ti	SF	ti	SF
0.23	0.998633	5.00	0.729727	8.07	0.525342	12.33	0.286644	23.20	0.023096
0.60	0.991424	5.07	0.724913	8.83	0.47805	13.27	0.24411	23.90	0.018662
0.77	0.986385	5.20	0.71598	8.87	0.475609	13.30	0.24282	26.03	0.009386
0.87	0.982986	5.30	0.709118	8.90	0.473781	13.80	0.221957	23.20	0.023096
1.83	0.938104	5.47	0.697473	8.93	0.471956	14.03	0.21276		
1.83	0.938104	5.50	0.695421	9.20	0.455661	14.33	0.201142		
2.03	0.926729	5.97	0.663419	9.23	0.453865	14.80	0.183802		
2.13	0.920858	5.97	0.663419	9.23	0.453865	15.10	0.173281		
2.30	0.910633	6.00	0.661386	9.30	0.449686	15.67	0.154456		
2.47	0.900136	6.10	0.654621	9.87	0.416283	15.83	0.149442		
2.50	0.898258	6.10	0.654621	9.93	0.412834	16.43	0.131678		
3.10	0.859461	6.20	0.647872	10.10	0.403133	16.57	0.127765		
3.10	0.859461	6.23	0.64585	10.17	0.399169	16.70	0.124209		
3.23	0.850811	6.73	0.61239	10.17	0.399169	16.87	0.119669		
3.33	0.844114	6.83	0.605755	10.36	0.388505	17.07	0.114488		
3.37	0.841426	6.93	0.59914	10.70	0.369772	17.33	0.108006		
3.43	0.837384	7.00	0.594522	10.73	0.36814	17.40	0.106309		
3.53	0.830625	7.07	0.589914	10.80	0.364349	18.03	0.091935		
3.67	0.82112	7.30	0.574851	10.93	0.35736	18.07	0.091075		
4.20	0.784828	7.43	0.56639	11.50	0.327544	20.27	0.052676		
4.40	0.771061	7.50	0.561851	11.87	0.308934	20.80	0.045746		
4.90	0.73661	7.57	0.557323	11.97	0.304008	21.17	0.041369		
5.00	0.729727	8.00	0.52978	12.03	0.301073	21.23	0.040693		

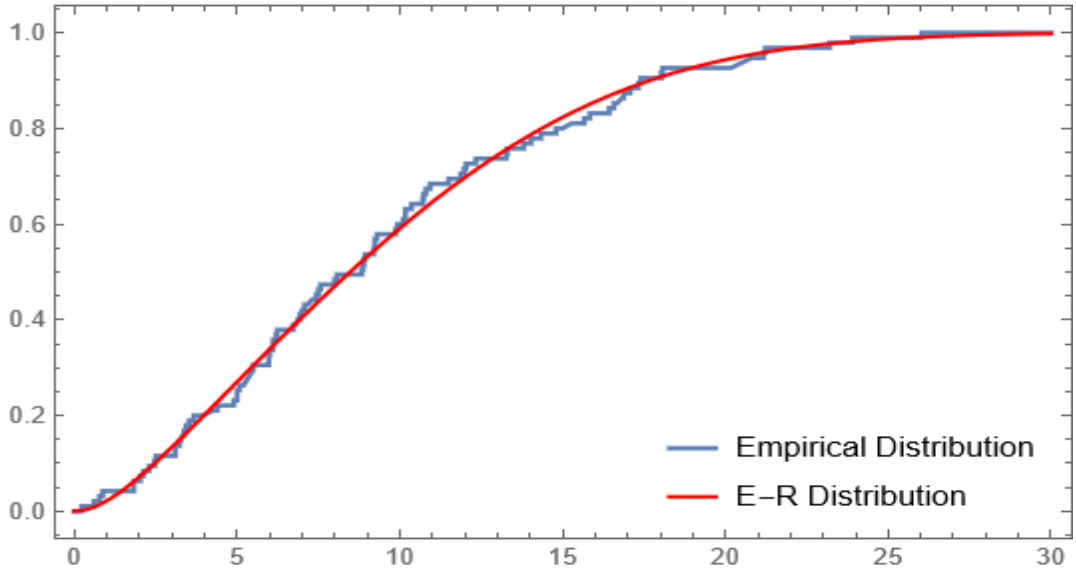
نلاحظ ان متوسط هذه القيم هو **0.499893** أي ان متوسط دالة البقاء هو 0.50 أي أن احتمال بقاء المريض بسرطان الثدي على قيد الحياة يبلغ 50%.

الشكل(2) دالة ال **CDF** للتوزيع المختلط للبيانات الحقيقية



اذ نلاحظ ان قيم **CDF** تقع بين الصفر والواحد وهذا ما يتفق مع النظرية الاحصائية

الشكل (3) دالة البقاء للتوزيع المختلط للبيانات الحقيقية



اذ نلاحظ من الشكل ان دالة البقاء تتناقص مع الوقت أي تتناسب عكسيا مع الوقت وهذ ما يتلاءم مع النظرية الإحصائية.

الاستنتاجات:

ان افضل توزيع في حسن المطابقة لبيانات البقاء كان التوزيع المختلط.

ان افضل توزيع في تقدير دالة البقاء كان التوزيع المختلط.

ان احتمال بقاء مريض سرطان الثدي على قيد الحياة هو 50%.

ان عملية خلط التوزيعات جعلت عملية تقدير دالة البقاء افضل من التوزيعات المفردة أي ان خلط التوزيعات يحسن من عملية التقدير.

دراسات مستقبلية :

1-نوصي باستعمال عدة طرق في تقدير دالة البقاء وللتوزيعات الثلاث والمقارنة بينها أيها افضل في التقدير.

2-نوصي باستعمال التضبيب في البيانات وتقدير دالة البقاء.

3-نوصي بالمقارنة بين أسلوب بيز والأسلوب الكلاسيكي في تقدير دالة البقاء.

المصادر:

1- السامرائي . حاتم (2022) " تقدير دالة المخاطرة للتوزيع المركب (جمبرتز – الأسي) مع تطبيق عملي" ، رسالة ماجستير مقدمة الى كلية الادارة والاقتصاد / جامعة كربلاء

2- السياب,كرار راضي منعم,(2023)," تقدير دالة البقاء التوزيع **Xgamma** مع تطبيق عملي", رسالة ماجستير في علوم الإحصاء,كلية الإدارة والاقتصاد,جامعة كربلاء.

3- القريشي ,كنعان عدنان احمد,(2022),"بناء توزيع احتمالي باستعمال قاعدة**(Topp Leone G-Famil)** مع تطبيق عملي", رسالة ماجستير في علوم الإحصاء,كلية الإدارة والاقتصاد,جامعة كربلاء.

4- عبدالحسين , غفران خالد , زعلان,ريسان عبدالامام,(2022)," تقدير معلمات ودالة البقاء لتوزيع مختلط مقترح باستعمال بعض طرائق التقدير من خلال المحاكاة",مجلة الاقتصادي الخليجي,العدد54.

5-**Ahmad, K. E., Jaheen, Z. F., & Mohammed, H. S. (2011). Finite mixture of Burr type XII distribution and its reciprocal: properties and applications.** ***Statistical Papers, 52,*** **835-845**.

6-**Fatima Bi, Afaq Ahmad, "Bayesian Estimation of the parameter of Ailamujia Distribution using different Loss functions" ,** ***International Journal of Research in Advent Technology, E-ISSN: 2321-9637,2019, India***

7-**Kahkashan Ateeq,& others ," An extension of Rayleigh distribution and Applications", Cogent Mathematics & Statistics(2019), 6:**1162219**https://doi.org/10.1080/25742558.2019.16 22191.**

8-**Naylor, T.H. & others (1966), "Computer Simulation Techniques",Willey**

9- **Smadl, Mahmoud M. Ansari , Saiful Islam," Allamujia Inverted Weibull distribution with application to lifetime data", Journal of Statistics , Vol. 38(3), 341-358.2022 Pakistan.**