توظيف اسلوب Jacknife لايجاد مقدرات معولية توزيع رايلي ذات المعلمة الواحدة ومقارنتها ببعض طرق التقدير الاخرى

أ.م.غفران اسماعيل كمال

جامعة بغداد/كلية الادارة و الاقتصاد قسم الاحصاء

المستخلص

يستعرض هذا البحث بعض طرائق تقدير معولية توزيع رايلي ذو المعلمة الواحدة وهي طريقة الأمكان الأعظم وطريقة العزوم فضلا عن طريقة المربعات الصغرى، وطريقة أنحدار Ridge، وطريقة أنحدار Ridge المعدلة، وطريقة جاك نايف . وتم أستعمال أسلوب تجارب المحاكاة بقيم معلمات مختلفة ولحجوم عينات مختلفة . لقياس كفاءة المقدرات اعتمادا على معيار احصائي ياخذ التباين والتحيز بعين الاعتبار هو معيار متوسط مربعات الخطأ MSE, و متوسط مربعات الخطأ التكاملي (IMSE) .

Abstract

This Paper reviewing some of estimation methods of Rayleigh distribution parameters which are: Maximum Likelihood, Moments, Least Square Method, Ridge Regression Method, Modified Ridge Method, Jackknife Method, Simulation experiments have been used with different parameter imposed values and for different sample sizes. Efficiency of the estimators has been measured by Mean Square Error (MSE) and (IMSE) .

المقدمة وهدف البحث

اخذت مسألة التقدير اهتماما" كبيرا" في التطبيقات الإحصائية والهندسية ومختلف العلوم التطبيقية والإنسانية لما تقدمه من وسائل ساعدت في التعرف بصورة اكثر دقة على العديد من العمليات المشوبة بأخطاء عشوائية , ولكن بعد التوسع الحاصل في استعمال موضوع التقدير في التوزيعات الاحصائية فقد ظهرت طرق مختلفة للتقدير معلمية ومنها لا معلمية , وفي بحثنا هذا سوف نتطرق الى مجموعة من الطرق لتقدير معلمة القياس Scale parameter بالنسبة الى توزيع رايلي والمقارنة بين هذه الطرق بالاعتماد على مقاييس احصائية لمعرفة الافضل منها تحت ظروف مختلفة هما معيار متوسط مربعات الخطأ MSE , و متوسط مربعات الخطأ التكاملي (IMSE) لذا يهدف البحث الى المقارنة بين طريقة Jacknife اللامعلمية وبعض الطرق المعلمية الاخرى بغية الوصول الى الطريقة المثلى للتقدير.

الجانب النظري

1- توزيع رايلي Rayleigh Distribution

يعد هذا التوزيع من التوزيعات الأحتمالية المهمة في مجال تحليل الخطأ سواءاً لمختلف الأنظمة أو في التحليلات المفردة. ويعد أنموذج للفشل في مجال أختبار الحياة. كذلك يستخدم توزيع رايلي في دراسة ظاهره ارتفاع الامواج البحرية في المحيطات وايضاً في دراسة سرعة الرياح , وكذلك من استخداماتة في معرفة قوة الاشارات اللاسلكية والراديوية في ساعات الذروة للاتصلات . وكالاتي:

فأذ أفترضنا أن: Y ~ Ray (θ) وبالدالة الأحتمالية:

$$f(t;\theta) = \frac{2}{\theta} \cdot t \cdot \exp\left\{-\frac{t^2}{\theta}\right\} \qquad ; \qquad I(t)_{(0,\infty)} \qquad \dots (1)$$

وهي ماتدعى دالة الكثافة الأحتمالية لتوزيع رايلي. أما الدالة التوزيعية المتراكمة له فتعطى بالصيغة:

$$F(t;\theta) = 1 - \exp\left\{-\frac{t^2}{\theta}\right\} \qquad \dots (2)$$

2- طرائق التقدير Estimation Methods

هنالك طرائق عديدة لتقدير معلمة النموذج الا أن التقليدية منها تفترض أن المعلمة المراد تقديرها ثابتة وليست متغيرة، وبذا يمكن تقديرها عندما تكون غير معلومة بأستعمال بيانات لمشاهدات العينة ومن اهم الطرق هي:

2-1- طريقة الأمكان الأعظم Maximum Likelihood Method

أن الخصائص الجيدة التي تمتاز بها هذه الطريقة جعلتها من الطرائق المهمة للتقدير. وتستند بالأساس الى ايجاد قيمة تقدير المعلمة التي تجعل دالة الأمكان عند نهايتها العظمى. ويمكن أعطاء دالة الأمكان على أنها دالة أحتمالية مشتركة بالصيغة الآتية:

$$L = f(t_1;\theta)\cdot f(t_2;\theta)\cdot f(t_3;\theta)\cdots f(t_n;\theta)$$

$$L=\left(\frac{2}{\theta}\right)^{n}\cdot\prod_{i=1}^{n}t_i\cdot\exp\left\{-\frac{\sum_{i=1}^{n}t_i^2}{\theta}\right\} \qquad \dots (3)$$

ولغرض تقدير دالة الأمكان يتم أخذ اللوغاريتم الطبيعي (ln) لطرفي المعادلة (3) أعلاه، فيتم الحصول على الصيغة الآتية:

$$\ln(L)=n\ln(2)-n\ln(\theta)+\sum_{i=1}^{n}\ln(t_i)-\frac{\sum_{i=1}^{n}t_i^2}{\theta} \qquad \dots (4)$$

ولغرض أيجاد القيمة التقديرية للمعلمة θ المجهولة والتي تجعل لوغاريتم دالة الأمكان في نهايته العظمى، يتم أخذ المشتقة الجزئية الأولى للمعادلة (4) أعلاه بالنسبة للمعلمة θ ومساواتها بالصفر ليتم الحصول على الآتي:

$$\hat{\theta}_{ML}=\frac{\sum_{i=1}^{n}t_i^2}{n} \qquad \dots (5)$$

2-2- طريقة العزوم Moment Method

وتعد هذه الطريقة من الطرائق التقليدية الشائعة لسهولتها. وتستند بالأساس الى مساواة عزم المجتمع μi بعزم العينة mi ، أي ان:

$$\mu_i(\theta_1,\theta_2,\theta_3,\cdots,\theta_k)=m_i(t_1,t_2,t_3,\cdots,t_k) \qquad \dots (6)$$

أذ أن i=1, 2, 3, …, k وأن k تمثل عدد المعلمات في النموذج الأحتمالي، وأن:

$$m_1(t_1,t_2,t_3,\cdots,t_k)=\frac{1}{n}\sum_{i=1}^{n}t_i=\bar{t} \qquad \dots (7)$$

ولتوزيع رايلي في المعادلة (1)، يتم أستخراج العزم الأول للمجتمع وهو مايدعى أحياناً بمعدل وقت لفشل Mean time to failure، وكما يأتي:

$$\mu_1=E(t)=\frac{\sqrt{\pi\theta}}{2} \qquad \dots (8)$$

ومن مساواة (7) و(8) يتم الحصول على:

$$\hat{\theta}_{MOM} = \frac{4}{\pi}(\bar{t})^2 \qquad \dots (9)$$

2-3 طريقة إنحدار Ridge **Ridge Regression Method(RR)**

وهي من الطرائق التي تعول على أيجاد مقدرات الأنحدار بالأستناد الى مصفوفة المعلومات X´X وكما يأتي:

$$\underline{\hat{b}}_{(RR)} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix} = \left(X'X + kI\right)^{-1} X'Y \qquad \dots (10)$$

إذ أن $0 < k < 1$ تمثل معامل Ridge

وأن I هي مصفوفة وحيدة Identity بأبعاد p×p

وأن p هو عدد المعلمات في نموذج الأنحدار وهي هنا (p = 2).

ويتم الحصول على تقديرات معلمة توزيع رايلي من جراء أستعمال تقديرات معلمتي أنحدار Ridge كما وردت في المعادلتين (11) و (12)، ولكن بأستبدال معلمتي الأنحدار المستخرجة بطريقة المربعات الصغرى بتلك المقدرة بأسلوب أنحدار Ridge.

2-4 طريقة أنحدار Ridge المعدلة (Modified Ridge Regression(MRR

وهي طريقة شبيهة بطريقة أنحدار Ridge إلاّ أن أختيار قيمة k ستستبدل بأخرى يعتقد أنها تؤثر في كفاءة تقدير معلمتي نموذج الأنحدار b0 و b1 والتي يتم الحصول على تقديراتها من خلال:

$$\underline{\hat{b}}_{(RRR)} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix} = \left(X'X + k^* I\right)^{-1} X'Y \qquad \dots (11)$$

أذ أن: k* يتم أيجادها بالصيغة الآتية:

$$k^* = \frac{pS_{ls}}{\underline{\hat{b}}_{LS}{}' \, \underline{\hat{b}}_{LS}} \qquad \dots (12)$$

وأن:

$$S_{LS} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{b}_0 - \hat{b}_1 X_i\right)^2}{n-p} \qquad \dots (13)$$

وبالأسلوب ذاته، يتم الحصول على تقديرات معلمة توزيع رايلي من جراء أستعمال تقديرات معلمتي أنحدار Ridge المعدلة كما وردت في المعادلتين (11) و (12)، ولكن بأستبدال معلمتي الأنحدار المستخرجة بطريقة المربعات الصغرى بتلك المقدرة بأسلوب أنحدار Ridge المعدلة.

2-5 طريقة جاك–نايف Jackknife Method

تعتمد هذه الطريقه على القيام بعمليات الحذف والارجاع للمفردات وبشكل متكرر ومتسلسل

للحصول على العينات الجزئية المتعددة والناتجة من الحذف، فبأفتراض حالة توفر عينة بحجم (n) من المفردات تسلك سلوك المتغير العشوائي ضمن توزيع رايلي، عندها يمكن تشكيل عدداً من العينات الجزئية (Partial Sample) وكل واحدة منها تتألف من عدد من المفردات بحيث ان :–

$$x_{ij} \quad ; \; i = 1, 2, 3, \dots, n-k \quad ; \quad j = 1, 2, 3, \dots, m$$

أذ أن: (k) تمثل عدد المفردات المحذوفة في كل مرة.

(n-k) تمثل حجم العينة الجزئية الناتج من بعد عملية الحذف.

(m) تمثل عدد العينات الجزئية الناتجة.

وبأستعمال العينات الجزئية يمكن الحصول على (m) من المقدرات الجديدة، والتي يمكن الحصول عليها من تطبيق المعادلة الآتية:

$$\hat{\theta}_{Jack} = n\hat{\theta}_{LS} - (n-1)\sum_{i=1}^{m} \frac{\hat{\theta}_{(\cdot)i}}{m} \qquad \dots (14)$$

أذ ان:

$\hat{\theta}_{Jack}$ تمثل المقدر على وفق طريقة جاك–نايف.

$\hat{\theta}_{LS}$ تمثل المقدر على وفق احدى طرائق التقدير الكلاسيكية، (وهنا تم أستعمال تقدير المربعات الصغرى).

3. الجانب التجريبي

3-1 المحاكاة

تم أعداد برنامجاً خاصاً بإستعمال لغة MATLAB version 7.4 (Release 13) البرمجية في أجراء تجارب المحاكاة بمراحلها من توليد البيانات الى أستخراج المقدرات وأخيراً أستخراج قيم معايير المفاضلة بين

الطرائق. أذ تم أعادة التجريب لـ (1000) تكرار، وللنماذج الثلاثة الآتية وبحجم عينة (n = 20, 35, 100) و (θ = 0.2 , 1.4 , 2) :

النموذج الأول: للدالة الأحتمالية . Rayleigh (θ = 0.2)

النموذج الثاني: للدالة الأحتمالية. Rayleigh (θ = 1.4)

النموذج الثالث: للدالة الأحتمالية. Rayleigh (θ = 2)

3-2 معايير المقارنة

3-2-1 متوسط مربعات الخطأ Mean Squared Error (MSE)

ويمثل معيار مقارنة تكون فيه الأفضلية للقيمة الأصغر، وكما في الصيغة:

$$MSE = \frac{1}{r}\sum_{i=1}^{r}\left(\theta_i - \hat{\theta}_i\right)^2 \qquad \ldots (15)$$

إذ ان θ : هي معلمة التوزيع.

r: هو عدد التكرارات في تجربة المحاكاة.

3-2-2 متوسط مربعات الخطأ التكاملي (IMSE)

هومقياس نسبي يستخدم لمقارنة دقة القياسات في حالة كون المقدرات مختلفة ويحتسب هذا المقياس وفق الصيغة الاتية

$$IMSE(\hat{R}(t)) = \frac{1}{r}\sum_{i=1}^{r}\left\{\frac{1}{n}\sum_{j=1}^{n_i}[\hat{R}_i(t_j) - R_j(t_j)]\right\}$$

3.3 النتائج

جدول (1) متوسط مربعات الخطأ MSE لمقدرات معلمة القياس θ

Model	Sample size	ML	MOM	Jack-Naif	RR	MRR	Best
1	20	0.004087	0.173193	0.003179	0.011681	0.003688	JN
	35	0.0022	0.228492	0.001343	0.004212	0.001469	JN
	100	0.000756	0.291311	0.000411	0.000806	0.000429	JN
2	20	0.003825	0.173016	0.002925	0.002765	0.003391	RR
	35	0.002089	0.224664	0.001269	0.001245	0.001389	RR
	100	0.000713	0.289477	0.000394	0.000396	0.000409	JN
3	20	0.004137	0.174077	0.002419	0.002191	0.002778	RR
	35	0.002305	0.222986	0.001341	0.001308	0.001412	RR
	100	0.000739	0.291141	0.000376	0.000377	0.00039	JN

نلاحظ تقدم واضح لطريقة JN خصوصا عند حجم عينة (n=100) تليها طريقة RR خصوصا عند حجم عينة (n=20,35) للنموذج الثاني والثالث

جدول (3) متوسط مربعات الخطأ التكاملي IMSE لمقدرات معلمة القياس θ

Model	Sample size	ML	MOM	Jack-Naif	RR	MRR	Best
1	20	0.000447	0.00853	0.000556	0.002324	0.000759	ML
	35	0.00025	0.01788	0.000173	0.000665	0.000198	JN
	100	9.06E-05	0.058966	5.16E-05	0.00011	5.45E-05	JN
2	20	0.001388	0.053909	0.001126	0.001074	0.001386	RR
	35	0.000776	0.071817	0.000494	0.000485	0.000551	RR
	100	0.000268	0.084193	0.000149	0.00015	0.000156	JN
3	20	0.001805	0.063946	0.00116	0.001049	0.001424	RR
	35	0.001014	0.087351	0.000596	0.000581	0.000637	RR
	100	0.000328	0.110143	0.00017	0.000171	0.000177	JN

استمر التميز لطريقة RR عند حجم عينة (n=20,35) للنموذج الثاتي والثالث يليه طريقة JN خصوصا عند حجم عينة (n=100) .

4. الأستنتاجات

يمكن أستخلاص بعض الأستنتاجات على ضوء نتائج تجارب المحاكاة الموصوفة في الفصل البحث ، وكمايأتي:

1. اظهرت طريقة JN تفوقا على طريقة RR عند تقدير معلمة القياس لتوزيع رايلي حيث كان متوسط مربعات الخطأ للمقدر وعند n كبيرة اصغر بكثير من (MSE) لمقدرات العزوم والامكان الاعظم .

2. كان هناك توافق (أو تقارب) في قياس الكفاءة (أو الأفضلية) لمابين المعيارين MSE و IMSE.

3- يمكن تطبيق البحث على بيانات حقيقية , وبيان اهمية هذا التوزيع من الناحية العملية .

المصادر

[1] A. Ahmed Soliman, "Comparison of Linux and Quadratic Bayes estimators for the Rayleigh Distribution". Communications in Statistics.-Theory and Methods, 29(1) (2000), 95-107.

[2] A. M. Abd Elfattah, Amal S. Hassan and D.M. Ziedan, "Efficiency of Maximum Likelihood Estimators under Different Censored Sampling Schemes for Rayleigh Distribution."
Interstat. March (2006a)

[3] A. M. Abd Elfattah, Amal S. Hassan and D.M. Ziedan. "Efficiency of Bayes estimators for Rayleigh distribution". Interstat, July (2006b

[4] A.M. Polovko, "Fundamental of Reliability Theory. Academic Press, New York, 1968.

[5] D.D. Dyer and C.W. Whisenand, "Best Linear Unbiased Estimator of the Parameter of the Rayleigh Distribution." IEEE Transactions on Reliability, R-22(1973), 27-34.

**[6] H.A. Howlader and A. Hossain," On Bayesian Estimation and Prediction from Rayleigh based on Type-II Censored Data. "Communications in Statistics.-Theory and Methods,
24(9) (1995), 2249-2259.**

**[7] J. A. Hartigan, "Invariant Prior Distribution. "Annals of .Mathematical Statistics, 34 (1964)
836-845**

**[8] Kundu, D. and Raqab, M.Z. (2005). "Generalized Rayleigh distribution": di®erent meth-
ods of estimation, *Computational Statistics and Data Analysis*, 49, 187-200.**