دراسة مقارنة
بعض طرائق تقدير معالم توزيع Rayleigh

احمد نعيم فليح / مدرس /كلية الادارة والاقتصاد /جامعة القادسية

الخلاصة :

يهتم هذا البحث بدراسة بعض طرائق التقدير لمعالم توزيع ريلاي (Rayleigh) حيث تم استخدام طريقة المربعات الصغرى الاعتيادية (OLS)، طريقة الانحدار المنتظم Regularized Regression (R.R)، طريقة الانحدار المنتظم الحصين Robust Regularized Regression وطريقة العزوم Method of moments of (M.O.M) لتقدير معلمة المكان Location Parameter ومعلمة القياس Scale Parameter لتوزيع (Rayleigh)، بعد ذلك المفاضلة بين تقديرات طرائق التقدير في أعلاه باستخدام معيار الانحرافات الكلية Total Deviation (T.D) و معيار متوسط مربعات الخطأ Mean Squared Error (M.S.E) ، حيث تمت المقارنة بين الطرائق التي في أعلاه باستخدام أسلوب المحاكاة (Simulation) .

المقدمة :

يعد توزيع Rayleigh من التوزيعات المهمة(2) في العديد من التطبيقات الفيزياوية والإلكترونية ونظرية العول (Reliability theory) ويعتبر هذا التوزيع حالة خاصة من توزيع (Weibull) حيث نجد ان دالة الكثافة الاحتمالية لتوزيع Weibull هي :

$$f(x,\gamma,\alpha,\beta)=\alpha\beta^{-\alpha}(x-\gamma)^{\alpha-1}\;.\;e^{-(x-\gamma)^{\alpha}/\beta^{\alpha}}$$

$$;x \geq \gamma \geq 0\,,\alpha > 0,\beta > 0\;.1) \qquad ....(1)$$

أي أن :

X ~ weibull (γ , α , β)

حيث أن γ هي معلمة المكان location parameter ، α هي معلمة الشكل shape parameter و β هي معلمة القياس scale parameter فإذا كانت α = 2 أي أن :

X ~ weibull (γ , 2 , β)

عندئذ يمكن أن نقول إن :

X ~ Raylieigh (γ , β)

و من العلاقة بين توزيع weibull و توزيع Normal[(2)و(6)] نجد انه إذا كان :

y ~ Normal (0 , β^2) ,

Z ~ Normal (0 , β^2)

علما" إن y متغير عشوائي مستقل عن المتغير Z فان :

$X = (Y^2 + z^2)^{1/2} \sim \text{weibull} (\gamma , \alpha =2 , 2^{1/2} \beta)$

أي أن :

$x \sim \text{Rayleigh} (\gamma , 2^{1/2} \beta)$

و بهذا فان دالة الكثافة الاحتمالية (p.d.f) لتوزيع Rayleigh يمكن أن تكون بالشكل التالي :

$$f(x;\gamma,\beta)=(x-\gamma)/\beta^2 \, . \, e^{-(x-\gamma)^2/2\beta^2} \quad ....(2)$$

حيث إن $\gamma < x < \infty$

و أن $\beta > 0$

كذلك فان الدالة التوزيعية (c.d.f) لتوزيع Rayliegh هي :

$$F(x)=1-e^{-(x-\gamma)^2/2\beta^2} \quad ....(3)$$

فقد وضح الباحث بالجانب النظري للبحث طرائق تقدير معلمة المكان(γ) Location parameter ومعلمة القياس (β) Scale parameter لتوزيع (Rayliegh) وهذه الطرائق هي طريقة المربعات الصغرى (OLS) وطريقة الانحدار المنتظم (R.R) وطريقة الانحدار المنتظم الحصين (R.R.R) وطريقة العزوم (M.O.M) ، كذلك تمت المفاضلة بين تقديرات المعالم لهذه الطرائق باستخدام أسلوب المحاكاة وبالاعتماد على معيار الانحرافات الكلية (T.D) ومعيار متوسط مربعات الخطأ (MSE) حيث ان [7] :

$$T.D=\left|\frac{\hat{\beta}-\beta}{\beta}\right|+\left|\frac{\hat{\gamma}-\gamma}{\gamma}\right| \quad ....(4)$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{F}(x_i)-F(x_i)\right]^2 \quad .....(5)$$

علما" ان :

$$\hat{F}(x_i)=1-e^{-(x_i-\hat{\gamma})^2/2\hat{\beta}^2}$$

$$F(x_i)=1-e^{-(x_i-\gamma)^2/2\beta^2}$$

الجانب النظرى :

س يتم في هذا الجانب توضيح طرائق تقدير معلمة الشكل (γ) و معلمة القياس (β)الاتية :-

1- طريقة المربعات الصغرى الاعتيادية

: Ordinary Least Squared Method (OLS)

تعد طريقة (OLS) من الطرائق الأكثر شيوعاً [1] لتقدير معالم نموذج الانحدار وذلك لسهولة وبساطة هذه الطريقة ومنطقية النتائج التي يتم الحصول عليها ، ان الهدف من طريقة المربعات الصغرى هو :

$$Minimize \sum_{i=1}^{n} r_i^2 \qquad ....(6)$$

حيث أن :

$$r_i = y_i - \hat{y}_i$$

و أن

$$\hat{y}_i = x_{i1}\,\hat{\beta}_1 + ... \, x_{ip}\,\hat{\beta}_p \qquad .....(7)$$

وبهدف استخدام هذه الطريقة لتقدير معلمة المكان (γ) ومعلمة القياس (β) لتوزيع Rayleigh نفترض بأن لدينا معادلة انحدار ، هذه المعادلة سيتم افتراضها بالاعتماد على الدالة التوزيعية (c.d.f) لتوزيع Rayleigh وكما موضح في الآتي :

$$F(x) = 1 - e^{-(x-\gamma)^2 / 2\beta^2}$$

$$1 - F(x) = e^{-(x-\gamma)^2 / 2\beta^2}$$

وبأخذ ال (ℓn) للطرفين نجد أن :

$$\ell n \left[1 - F(x)\right] = -(x-\gamma)^2 / 2\beta^2$$

وبهذا نجد أن :

$$(x-\gamma)^2 = -2\beta^2 \ell n[1-F(x)]$$

من المعادلة الأخيرة نجد أن :

$$x-\gamma = \beta\{-2\ \ell n[1-F(x)]\}^{1/2}$$

أي أن :

$$x = \gamma + \beta\{-2\ \ell n[1-F(x)]\}^{1/2} \ldots\ldots (8)$$

إن المعادلة رقم (4) يمكن اعتبارها معادلة انحدار فيها المتغير المعتمد هو (x) والمتغير المستقل هو $\{-2\ \ell n[1-F(x)]\}^{1/2}$

حدها الثابت هو المعلمة (γ) وميلها هو المعلمة (β) ومن ثم فان القيم المقدرة حسب طريقة (OLS) للمعالم (γ) و (β) يمكن إيجادها باستخدام الصيغ التقديرية الآتية :

$$\hat{\gamma} = \frac{\sum_{i=1}^{n} y^2{}_i \sum_{i=1}^{n} x_i - (\sum y_i).(\sum x_i y_i)}{n(\sum y^2{}_i) - (\sum y_i)^2} \qquad \ldots\ldots\ldots(9)$$

$$\hat{\beta} = \frac{n\sum x_i y_i - (\sum y_i)(\sum x_i)}{n\sum y_i{}^2 - (\sum y_i)^2} \qquad \ldots\ldots.(10)$$

حيث أن :

$$y_i = \{-2\ell n[1-F(x_i)]\}^{1/2}$$

2- طريقة الانحدار المنتظم Regularized Regression (R.R)

وهي تعميم طفيف للانحدار الخطي [5] تعتبر طريقة الانحدار المنتظم أو ما يسمى بـ (Ridge Regression) إحدى طرائق تقدير معالم نموذج الانحدار وتعتبر هذه الطريقة في بعض الأحيان أدق من طريقة (OLS) وذلك عندما يعاني نموذج الانحدار من مشكلة التعدد الخطي ، حيث اقترح ( ) صيغة رياضية يمكن من خلالها أيجاد مقدرات الانحدار المنتظم وكالآتي :

$$\hat{\beta}_{R.R} = (x^T x + CI)^{-1} \,.\, x^T Y \quad \ldots\ldots (11)$$

حيث أن C هو معامل الانحدار المنتظم ، 0<C<1

و I هي مصفوفة الوحدة ذات المرتبة pxp حيث أن p هو عدد المعالم المقدرة .

تعرف الصيغة (11) بمقدر انحدار Regularized الاعتيادي (O.R.R) علماً أن أحد أساليب تقدير قيمة (C) هو أسلوب المحاكاة وباستخدام الوسط التوافقي لقيم (C_i) يهدف إعطاء قيمة مفردة لـ (C) نجد ان قيمة (C) يمكن حسابها من المعادلة الاتية :

$$C = \frac{P\,S_e^2}{\beta'_{OLS} \cdot \beta_{OLS}} \quad \ldots\ldots\ldots\ldots (12)$$

حيث أن P هو عدد المعالم المطلوب تقديرها ، وأن

$$S_e^{\,2} = \frac{\sum_{i=1}^{n} e_i^{\,2}}{n-p} \quad \ldots\ldots\ldots\ldots\ldots (13)$$

3- طريقة الانحدار المنتظم الحصين

: Robust Regularized Regression (R.R.R)

إن الانحدار الحصين (Robust Regression) يحاول [3] خلق مقدرات تكون غير متأثرة جداً بالقيم الشاذة ، حيث إن معادلة الانحدار الاعتيادية عادةً ما تواجه مشكلة خطيرة أثناء تقدير المعالم بطريقة المربعات الصغرى إلا وهي وجود القيم الشاذة ، إن مقدر طريقة (R.R.R) يمكن كتابته كالآتي :

$$\hat{\beta}_{R.R} = (X^T X + K\, I)^{-1} X^T\, Y \quad \ldots\ldots\ldots\ldots (14)$$

علماً أن قيمة K يمكن تحديدها من البيانات كالآتي :

$$K = \frac{pS_R^2}{\hat{\beta}_R^T \,.\, \hat{\beta}_R}$$

و أن :

$$S_R^2 = \frac{\sum_{i=1}^{n} e_R^2}{n-p}$$

هذا يعني ان المقدر المحسوب بهذه الطريقة هو مقدر مركب من خصائص (R- estimators) مع (Regularized regression estimators).

لذلك يرمز له بالرمز $\hat{\beta}_{R.R}$. و ان الدوال S_R^2 ، $\hat{\beta}_R^T$ و e_R^2 يتم احتسابها بالاعتماد على طريقة (R-estimators) .

علما" ان (R-estimators) هي طريقة من طرائق تقدير المقدرات الحصينة لنموذج الانحدار () تعتمد أساساً على رتب البواقي (Ranks of residuals) فإذا فرضنا ان R_I هي رتبة الــ γ_i حيث أن :

$$r_i = y_i - x_i \beta$$

فان الهدف سيكون هو :

$$\underset{\hat{\beta}}{Minimize} \sum_{i=1}^{n} a_n (R_i).r_i \qquad ....$$

حيث أن $a_n(i)$ دالة رتيبة يجب أن تحقق الشرط الآتي :

$$\sum_{i=1}^{n} a_n(i) = 0 \qquad ....(15)$$

علماً أن الدالة $a_n(i)$ يمكن تقديرها بعدة طرائق وسنعتمد في بحثنا هذا على طريقة (Wilcoxon) الآتية :

$$a_n(i) = i - (n+1)/2 \qquad ....(16)$$

4 - طريقة العزوم Method of Moments (MOM) :

تعتبر طريقة العزوم (MOM) إحدى طرائق تقدير المعالم فعلى فرض انه لدينا عينة عشوائية حجمها n ومفرداتها هي $X_1, X_2, ...X_n$ فأن :

$$M'_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j \qquad \cdots\cdots(17)$$

حيث أن M'_j هو العزم وهو بنفس الوقت المقدر غير المتحيز .

فبالنسبة لتوزيع Rayleigh نجد أن العزم الأول (7) يمكن إيجاده بالمعادلة الآتية :

$$M'_1 = \sqrt{2}\ \beta\ \Gamma(\frac{3}{2}) + \gamma \qquad ....(18)$$

وأن العزم الثاني لتوزيع Rayleigh هو :

$$M'_2 = 2\beta^2 + \gamma\left[2\sqrt{2}\ \Gamma(\frac{3}{2})\beta + \gamma\right] \qquad \cdots\cdots(19)$$

وكذلك نجد أن :

$$\sigma^2 = M'_2 - M^2$$

$$\sigma^2 = \beta^2(2 - \frac{\pi}{2}) \qquad ....(20)$$

من المعادلة (20) نجد قيمة $\hat{\beta}$ أي أن :

$$\hat{\beta} = S\,\frac{\sqrt{2}}{\sqrt{4-\pi}} \qquad ....(21)$$

حيث أن S هو الانحراف المعياري للعينة .

ومن المعادلة (18) نجد قيمة $\hat{\gamma}$ بعد تعويض قيمة $\hat{\beta}$

أي أن :

$$\hat{\gamma} = m_1 - \sqrt{2}\hat{\beta}\,\Gamma(\frac{3}{2}) \qquad ....(22)$$

حيث أن m_1 هو عزم العينة .

الجانب العملي :

اعتمد الباحث في المفاضلة بين طرائق التقدير الذكورة سابقاً على أسلوب المحاكاة في تقدير المعلمة (γ) والمعلمة (β) ، حيث قام الباحث بتوليد عينات عشوائية بمعالم معلومة و لكل عينة تم تغيير حجم العينة ليكون (20,30,40) وتم استخدام معيار الانحرافات الكلية (T.D) ومعيار متوسط مربعات الخطأ (MSE) لكي تتمكن من المقارنة أو المفاضلة بين الطرائق المختلفة .

والجدول رقم (1) يبين النتائج التي تم الحصول عليها بالاعتماد على أسلوب المحاكاة حيث تم كتابة برنامج المحاكاة اعتماداً على لغة Q.Basic .

جدول رقم (1)

n	β	γ	حجم العينة	OLS				R.R			
				$\hat{\beta}$	$\hat{\gamma}$	T.D	MSE	$\hat{\beta}$	$\hat{\gamma}$	T.D	MSE
1	1	1	10 20 30	1 1 0.999992	0.999 0.9999 0.99991	0.0001 $2*10^{-6}$ $16*10^{-5}$	$1*10^{-6}$ $1*10^{-15}$ $64*10^{-14}$	1 1.005 0.998	0.9999 0.984 1.001	0.001 0.006 0.003	$2*10^{-7}$ $6*10^{-6}$ $45*10^{-9}$
2	2	2	10 20 30	2.00129 2.0045 2.0244	1.9976 1.9926 1.93346	0.0018 0.00595 0.048222	$1*10^{-7}$ $6*10^{-11}$ $74*10^{-12}$	2.004 2.014 1.93	1.955 1.961 2.031	0.0245 0.0265 0.0505	$9*10^{-9}$ $34*10^{-6}$ $16*10^{-7}$
3	3	3	10 20 30	3.0012 3.0048 3.0007	2.999 2.99134 3.0005	$73*10^{-4}$ $44*10^{-3}$ $39*10^{-4}$	0.0054 $42*10^{-10}$ $33*10^{-10}$	2.947 2.947 2.9993	3.001 3.014 2.9997	$18*10^{-2}$ $21*10^{-2}$ $33*10^{-4}$	0.041 $6*10^{-5}$ $32*10^{-11}$

n	β	γ	حجم العينة	RRR				MOM			
				$\hat{\beta}$	$\hat{\gamma}$	T.D	MSE	$\hat{\beta}$	$\hat{\gamma}$	T.D	MSE
1	1	1	10	1.0001	1.00003	0.00013	$46*10^{-10}$	0.8462	0.988	0.166	$93*10^{-8}$
			20	1.381	0.899	0.491	0.01715	0.8694	1.1021	1.014	0.29462
			30	1.00001	0.99999	0.00002	$38*10^{-13}$	0.9022	1.0673	0.1657	$73*10^{-6}$
2	2	2	10	2.001	1.998	0.0015	$56*10^{-10}$	1.692	1.9959	0.1665	$84*10^{-7}$
			20	2.005	1.9932	0.006	$38*10^{-9}$	1.7428	2.1973	0.2273	$11*10^{-9}$
			30	1.9335	2.02991	$48*10^{-8}$	$14*10^{-12}$	1.8058	2.133	0.1636	$72*10^{-9}$
3	3	3	10	3.001	2.9986	$80*10^{-4}$	$64*10^{-12}$	2.5397	2.634	$28*10^{-1}$	$48*10^{-8}$
			20	3.005	2.991	$47*10^{-3}$	$43*10^{-9}$	2.6124	3.298	$23*10^{-1}$	$11*10^{-12}$
			30	2.005	3.007	$41*10^{-3}$	$32*10^{-10}$	2.8605	3.2811	$14*10^{-1}$	$51*10^{-9}$

الاستنتاجات :

من كل ما تقدم نستنتج ما يأتي :

1- ان طريقة (RRR) تعطي مقدرات قريبة الى القيم الحقيقية المفترضة . و بعدها طريقة (OLS) و من بعدها طريقة (R.R) و من ثم طريقة MOM .

2- نجد ان قيم معيار (T.D) و معيار (MSE) كانت قيمها قليلة جدا" لطريقة (RRR) و من ثم طريقة (OLS) مقارنة" مع طريقة (RR) و طريقة MOM .

3- تعتبر طريقة العزوم (MOM) من اسهل و ابسط الطرق عند ايجاد قيمها حسابيا" من بيانات العينة لكن قيم معيار (T.D) و MSE كبيرا" جدا" مقارنة" مع نظيراتها في الطرائق الاخرى .

المصادر :

1- الشاروط ، محمد حبيب و مهند فائز ، 2000 " التقدير بطريقتي الحرف Ridge و المربعات الصغرى لدالة الانتاج في الاقتصاد العراقي للفترة 1999- 1970 " . مجلة وقائع المؤتمر الثاني عشر للجمعية العراقية الاحصائية .

2- ذنون ، باسل يونس ، 1991 " الاحتمالية و المتغيرات العشوائية" وزارة التعليم العالي و البحث العلمي ، جامعة الموصل ، مطبعة جامعة الموصل .

3- Heikkila ; Janne , (2003) " Robust regression " . department of electrical engineering . University of Oulu .

4- Horel , A.E and Kennard , R.W. (1970) , " Ridge regression Biased Estimation for Nonorthogonal problems " . Technometrics , 12, pp . 55-66 .

5- Jordan , I . Micheal , (2000) " linear and Ridge regression and Kernals " ; advanced Topics in learining and making decision .

6- Law , Averill M. and Ketton W.David , (1991) " Simulation modeling and analysis " , 2nd ed. , MCGraw-Hill series .

7- Mohamed A.Al Fawzan (2000) " Algorithms for estimating the parameters of weibull distribution " . Interstat , oct.2000 .