مقارنة بين طريقتي white التقليدية والحصينة لدالة و معولية توزيع رايلي ذي المعلمتين باستخدام المحاكاة

د. فاضل عبد العباس
استاذ مساعد
هيئة التعليم التقني

مجتبى زهير علي
طالب ماجستير رياضيات
جامعة الكوفة

الخلاصة:

يهتم هذا البحث بمسألة تقدير دالة المعولية لتوزيع رايلي بمعلمتين ، من خلال المقارنة بين طريقتي وايت التقليدية و الحصينة تم استخدام الدالة البرمجية robustfit في حساب تقديرات معلمتي توزيع رايلي ، كما تم اجراء دراسة تجريبية باستخدام المحاكاة لغرض المقارنة و اثبات كفاءة تلك الطرائق عمليا و ذلك من خلال الاعتماد على مشاهدات مولدة من توزيع رايلي بمعلمتين و لأحجام عينات و قيم معلمات مختلفة تمت المقارنة بالاعتماد على عدة مقاييس متمثلة بمتوسط مربعات الخطأ (MSE) و متوسط مربعات الخطأ للمقدرات (MSEθ) و معامل التحديد (R^2) . تبين ان طريقة وايت الحصينة هي أكفاءه طريقة و لجميع احجام العينات و قيم المعلمات المستخدمة في الدراسة .

ABSTRACT

This paper is concerned with estimation of reliability function of the two parameter Rayleigh distribution, through the comparison between the White traditional and Robust methods . we use the function code robustfit in calculating the estimates of the parameter of Rayleigh distribution was also a study of experimental use of simulation for the purpose of comparison and prove the efficiency of those methods in practice and that by relying on the views generated from the distribution of Rayleigh parameter and sample sizes and values of different parameters and were compared based on several measures of mean squares error (MSE) and the mean squares error of Estimates (MSEθ) and the coefficient of determination (R^2). And turned out that White Robust method is efficient method and for all sample sizes and values of the parameters used in the study.

1- المقدمة :

يعتبر توزيع رايلي من التوزيعات الاحتمالية المهمة في تحليل اوقات البقاء لكثير من الظواهر، كما ان توزيع رايلي هو انموذج للفشل واكتشف من قبل العالم الفيزيائي الانكليزي (Lord Rayleigh) ويستخدم في بحوث اختبار الحياة وفي التحليلات المفردة وتحليلات الخطأ لمختلف الانظمة، وكذلك وصف توزيع الوقت المستغرق لحين حصول الفشل في الوحدات المنتجة ممايساهم ذلك في توفير توقعات جيدة لاوقات الفشل المحتملة وخصوصا في حالة العينات الصغيرة [8] .

سوف نتناول في هذا البحث طريقتي وايت (white method) التقليدية والحصينة لتقدير معلمتي توزيع رايلي (α,β) و من ثم اعتماد هذه المقدرات في حساب دالة المعولية لتوزيع رايلي و ان المقارنة بين طريقتي وايت التقليدية و الحصينة ستتم باستخدام اسلوب المحاكاة من

خلال اعتماد متوسط مربعات الخطا (MSE) و MSE_{θ} و R^2 كاساس في المقارنة .

2- هدف البحث : يهدف البحث الى المقارنة بين طريقتي وايت التقليدية والحصينة لمعلمتي توزيع رايلي (α , β) حيث ان α تمثل معلمة الازاحة و β معلمة القياس وسوف نقوم بتقديرهما معا في حين اغلب البحوث تعتمد احدى المعلمات معلومة وثم تقدير المعلمة الثانية.

3- الجانب النظري[2]

ان دالة الكثافة الاحتمالية (p.d.f) لتوزيع رايلي ذي المعلمتين (α,β) والذي يعتبر حالة خاصة من توزيع وبيل عندما $(\gamma = 2)$ تاخذ الصيغة الاتية :

$$f(t;\alpha,\beta) = \begin{cases} \frac{2(t-\alpha)}{\beta}\ e^{\frac{-(t-\alpha)^2}{\beta}} & \alpha < t < \infty \,; \beta > 0 \\ 0 & o.w \end{cases} \quad \dots(1)$$

اذ ان:

α : تمثل معلمة الازاحة (shifting parameter)

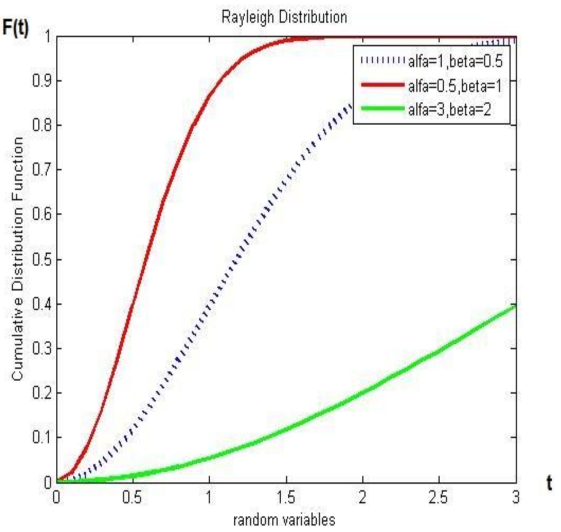
β: معلمة القياس (scale parameter)

ان توزيع رايلي بمعلمتين ممكن تطبيقة على المكائن و المعدات ذات معدل فشل متغير مع الزمن ، وايضا اذا كان وقت الفشل يبدا من زمن معين (α) و ليس من الصفر ، اذ ان (α) تمثل مدة العمر المضمون و تستخدم هذه المعلمة لوصف مدة الضمان او المدة التي لا تحدث فيها اعطال ابتدائية .

اما دالة التوزيع التراكمية (c.d.f) لتوزيع رايلي بمعلمتين (α,β) يمكن كتابتها حسب الصيغة الاتية :

$$F(t;\alpha,\beta) = p_r(T \le t) = \int_0^t f(u)du = 1 - e^{\frac{-(t-\alpha)^2}{\beta}} \quad \dots(2)$$

ويمكن توضيح شكل دالة التوزيع بيانيا حسب الشكل الاتي



شكل رقم (1) : دالة التوزيع التراكمية لتوزيع رايلي (c.d.f)

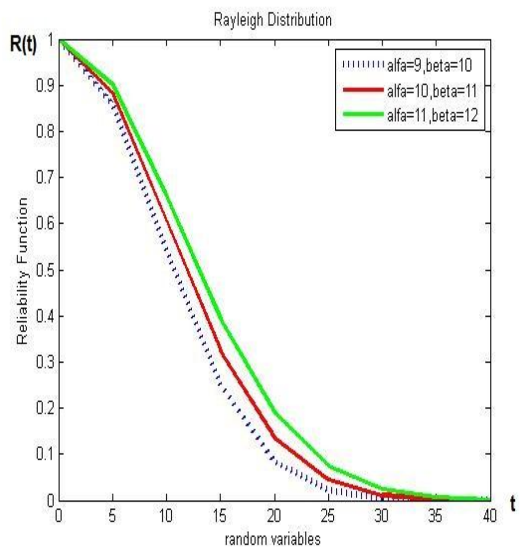
و دالة المعولية له فهي مكملة لدالة التوزيع التراكمية أي ان :

$$R(t) = p_r(T > t) = 1 - F(t) = \int_t^{\infty} f(u;\alpha,\beta)du = \int_t^{\infty} \frac{2(u-\alpha)}{\beta} e^{\frac{-(u-\alpha)^2}{\beta}} du$$

وبعد التكامل نحصل على

$$R(t) = e^{\frac{-(t-\alpha)^2}{\beta}} \quad \alpha < t < \beta \,, \beta > 0 \quad \dots(3)$$

و يمكن توضيح دالة المعولية حسب الشكل الاتي :-



شكل رقم (2): دالة المعولية بيانيا لتوزيع رايلي

وطبقا لذلك تكون دالة المخاطرة لهذا التوزيع وهي تمثل معدل الفشل $h(t)$ حسب الصيغة الاتية :-

$$h(t)=\frac{f(t)}{R(t)}=\frac{2(t-\alpha)}{\beta} \qquad ...(4)$$

نلاحظ من المعادلة (4) كيف ان معدل الفشل سيكون متغير مع الزمن

طرائق تقدير معلمات توزيع رايلي:

توجد العديد من طرائق التقدير لمعلمات توزيع رايلي ذي المعلمتين (α,β) و سوف نتناول منها التالي :-

1− طريقة وايت (White Method) [5]

تعتمد هذه الطريقة بصورة اساسية على دالة (c.d.f) المبنية في المعادلة (2) في صياغة انموذج انحدار الخطي البسيط و كما ياتي :-

$$F(t)=1-e^{-\frac{(t-\alpha)^2}{\beta}}$$

$$1-\mathrm{F}(t)=\mathrm{e}^{-(t-\alpha)^2/\beta}$$

باخذ اللوغارتم الطبيعي للطرفين و بضرب بالسالب

$$\ln[1-F(t)]^{-1}=\frac{(t-\alpha)^2}{\beta}$$

و باخذ اللوغارتم مرة اخرى نحصل على

$$\ln[\ln[1-\mathrm{F}(t)]^{-1}]=2\ln(t-\alpha)-ln\beta$$

أي تم توصل الى انموذج الانحدار الخطي الاتي :-

$$y_i=a+bx_i+\epsilon_i \ \forall\, i=1,2,\dots,n \ \dots(5)$$

حيث ان:-

$$y_t=\ln[\ln[1-\mathrm{F}(t)]^{-1}],\quad b=2 \qquad ,$$

$$a=-ln\beta,\ x_t=\ln(t-\alpha)$$

ϵ_i تمثل متغير الخطا العشوائي

و بتطبيق طريقة المربعات الصغرى على المعادلة اعلاه (5) نتوصل الى المقدرات الاتية:-

$$\hat{b}_{lsq}=\frac{\sum_{i=1}^{n}x_iy_i-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n}}{\sum_{i=1}^{n}{x_i}^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}}$$

$$\hat{a}_{\mathrm{lsq}}=\bar{\mathrm{y}}-\hat{\mathrm{b}}\bar{\mathrm{x}}$$

و يمكن الحصول على مقدرات معلمتي دالة رايلي($\hat{\beta}$ $\hat{\alpha}$,)في صيغة (1) وكما ياتي :-

$$\hat{\alpha}=\hat{\mathrm{b}}_{\mathrm{lsq}} \qquad \dots(6)$$

$$\hat{\beta}=e^{-\hat{a}_{lsq}} \qquad \dots(7)$$

و عليه فان مقدر وايت للدالة المعولية التقريبية يكون كماياتي :-

2- طريقة وايت الحصينة Robust White (Method) [6] [9]

ان وجود قيمة شاذة واحدة فقط قد تفقد المقدرات التقليدية مزاياها الجيدة ، بمعنى ان وجود قيمة شاذة واحدة ضمن مجموعة المشاهدات قد يؤ ثر في بعض مقدرات المعلمات حسب الطرائق التقليدية ،اذ تكون هذه المقدرات الى حد ما حساسة لمثل هذه المشاهدات مما قد يجعلها اقل كفاءة لما هي عليه عند عدم وجودها و ربما تفقدها الكثير من الخصائص الجيدة التي تتمتع بها وفي ضوء ذلك ظهرت الحاجة الى ابتداع مقدرات او اساليب اكثر

موضوعية تنطوي جميعها تحت مسمى الحصانة (Robustness).

فقد تم تطوير العديد من الطرائق الحصينة لتقدير معلمات النماذج الخطية منها طريقة M(تقديرات الأرجحية العظمى) وهي أكثر الطرائق شيوعاً وطريقة تقديرات Lρ المعتمدة على أساس التقريب لدالة Laplace وطريقة تقديرات R المعتمدة على تقدير الرتب. وجميع هذه الطرائق تتفق في خطوات أساسية كإعطاء وزن للمشاهدة الشاذة لتقليل تأثيرها مقارنة مع الأوزان التي تعطى لبقية المشاهدات والاعتماد على التكرار في عملية حساب المقدرات والتوقف عندما يصبح الفرق بين التقديرات المتعاقبة قليلاً.

و نتيجة التطور السريع في البرامجيات الحديثة ومنها تطبيق الماتلاب ظهرت دوال أحصائية ضمن هذه البرامجيات تستخدم كأسلوب لحل المشاكل التي تعاني منها النماذج الأحصائية ومن هذه الدوال دالتي Robust regression ودالة Robustfit ومن اجل الحصول على المقدرات اكثر كفاءة سوف نعتمد على الأنحدار الحصين من خلال عدة أساليب منها الدالة البرمجية robustfit في حساب تقديرات معلمتي توزيع رايلي على العكس في الطريقة السابقة حيث تم الاعتماد على طريقة المربعات الصغرى .

حيث ان الدالة البرمجية robustfit تكتب بالصيغة العامة التالية:-

b=robustfit(x,y,'wfun',tune,'const) …(9)

حيث ان :-

Y,X : متجه المتغيرات في أنموذج الأنحدارالخطي البسيط

wfun : هي عبارة عن دوال موزونة مقترحة من قبل الاحصائين و الباحثين باسلوب التقديرات الحصينة بحيث تعطي مقدر غير شديد الحساسية وغير متأثر بالشواذ .

أما *tune, const* فهي تمثل ثابت القطع

ولغرض الوصول الى المقدر الافضل فقد تم الاعتماد على المقاييس الاحصائية التالية كاساس للمقارنة :-

1- متوسط مربعات الخطأ (MSE) والذي ياخذ الصيغة الاتية

$$MSE\left(f(t_i)\right) = \frac{\sum_{i=1}^{n}\left((\hat{f}(t_i) - f(t_i)\right)^2}{n} \quad \dots (10)$$

حيث ان

$$f(t_i) = \frac{2(t-\alpha)}{\beta} e^{-(t-\alpha)^2/\beta}$$

2- مقياس معامل التحديد (R^2): الذي ياخذ الصيغة التالية

$$R^2 = 1 - \frac{TSS}{RSS} \qquad \dots (11)$$

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3- الاعتماد على ايجاد متوسط مربعات للمقدرات من خلال الصيغة التالية [3] :

$$MSE_\alpha = \frac{1}{rep}\sum_{i=1}^{rep} (\alpha_0 - \hat{\alpha})^2$$

$$MSE_\beta = \frac{1}{rep}\sum_{i=1}^{rep} (\beta_0 - \hat{\beta})^2$$

$$MSE_\theta = det \frac{1}{rep}\left(\sum_{i=1}^{rep} \left[\begin{bmatrix}\hat{\beta} \\ \hat{\alpha}\end{bmatrix} - \begin{bmatrix}\beta_0 \\ \alpha_0\end{bmatrix}\right] * \left[\begin{bmatrix}\hat{\beta} \\ \hat{\alpha}\end{bmatrix} - \begin{bmatrix}\beta_0 \\ \alpha_0\end{bmatrix}\right]'\right) \quad \dots (12)$$

4 - الجانب التجريبي [4] [7]

يمكن تعريف المحاكاة بأنها عملية تمثيل أو تقليد للواقع الحقيقي باستخدام نماذج معينة وبتعبير آخر يمكن أن تُعرّف المحاكاة على أنها "عملية إعطاء صورة بديلة لنظام من دون إعطاء صورة ذلك النظام، الذي هو عبارة

عن علاقة دالية تشمل مجموعة من الأجزاء التي تعرف بأنها مكونات ذلك النظام"[1].

تم اجراء الدراسة باستخدام اسلوب المحاكاة في توليد البيانات انموذج رايلي وبقيم افتراضية لمعلمة $\alpha = 1.5$ و المعلمة $\beta = 1$ و بالاعتماد على احجام العينات التالية (n=10,50,100) وبتكرار التجربة (L=1000) ، و باستخدام تطبيق البرنامج Matlab.

حيث تم توليد البيانات من التوزيع المنتظم وتحويلها الى بيانات توزيع رايلي ذو المعلمتين وباستخدام دالة التوزيع التجميعية وحسب طريقة التحويل العكسية التالية:

$$F(t,\alpha,\beta) = 1 - e^{-(t-\alpha)^2/\beta}$$
$$U = 1 - e^{-(t-\alpha)^2/\beta}$$
$$e^{-(t-\alpha)^2/\beta} = 1 - U$$
$$\frac{-(t-\alpha)^2}{\beta} = \ln(1-U)$$
$$(t-\alpha)^2 = -\beta\ln(1-U)$$
$$\therefore t = \alpha + \sqrt{-\beta\ln(1-U)} \qquad \ldots (13)$$

خطوات تجربة المحاكاة :

تتلخص تجارب المحاكاة في الخطوات التالية :

1- اختيار البرنامج الذي يمكننا التعامل معا و كذلك لدينا معلومات كافية لبدء المحاكاة ، في هذه الدراسة استخدمنا برنامج Matlab و ذلك بسبب خصائصه و مرونته مقارنة مع اللغات البرمجية الاخرى .

2- تحديد القيم الافتراضية لمعلمات توزيع رايلي التي سوف تستخدم لتوليد البيانات العشوائية ، و قد استخدمنا α=1.5 و $\beta = 1$.

3 - توليد بيانات عشوائية من خلال كتابة المعادلة التالية داخل برنامج الماتلاب وفقا لصيغة الاتية

```
t=alpha+(-bata*log(1-U)).^0.5 … (14)
```

حيث ان

```
U=rand(n,1)
```

4- تحديد احجام العينات من اهم المراحل التي تعتمد عليها مراحل لاحقة ، في هذه الدراسة استخدمنا حجم العينة على النحو التالي N=10 , 50 , 100 .

5- اختيار عدد معين لتمثيل تكرار التجربة و السيطرة على الاختلافات بين العينات العشوائية اخترنا r=1000 .

6- اختيار معدل التلوث من البيانات الطبيعية و قد افترضنا معدل التلوث في هذه الدراسة ليكون 20%

7- توليد بيانات ملوثة بنسبة 20% مع افتراض القيم للمعلمات كما يلي: $\alpha2 = 10$, $\alpha1$=1.5 و $\beta2 = 8$, $\beta1 = 1$.

حيث ان

u_1=rand(n_1 ,1)

u_2=rand(n_2 ,1)

$$t1 = \alpha1 + \sqrt{-\beta1\ln(1-u1)}$$
$$t2 = \alpha2 + \sqrt{-\beta2\ln(1-u2)}$$

عندما

N	$\widehat{\alpha}$	$\widehat{\beta}$	MSE	MSEθ	R^2
10	2.0027	1.3807	0.2980	2.9038e-004	0.9567
50	2.0002	1.3389	0.2872	1.6762e-006	0.9515
100	1.9985	1.3408	0.2855	2.1885e-007	0.9509

n_1 : تمثل حجم العينة للبيانات الطبيعية

n_2 : تمثل حجم العينة للبيانات الملوثة

و بالتالي نحصل على متجه للبيانات الملوثة

t=[t_1;t_2] … (15)

اولا:- الحالة الطبيعية للبيانات

سوف نستخدم المعادلة (14) لتوليد البيانات الطبيعية و الجداول التالية توضح حساب مقدرات معلمات توزيع رايلي

الجدول (1) يوضح تقديرات معالم توزيع رايلي حسب طريقة التقليدية عندما $\beta = 1 , \alpha = 1.5$

الجدول (2) : يوضح تقديرات معالم توزيع رايلي حسب طريقة الحصينة عندما $\beta = 1 , \alpha = 1.5$

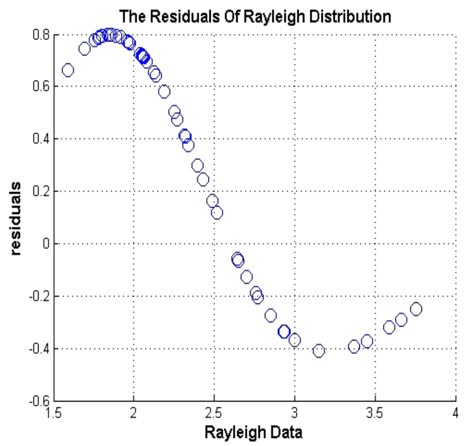
N	$\hat{\alpha}$	$\hat{\beta}$	MSE	MSEθ	R^2
10	1.0071	1.3806	0.0667	2.7260e-004	0.9531
50	1.0037	1.3388	0.0602	1.5582e-006	0.9486
100	1.0012	1.3407	0.0596	1.9007e-007	0.9480

من خلال تجارب المحاكاة في الجداول (1) و (2) فقد سجلت نتائج مؤشرات عمليات المقارنة ما بين الطريقتين و المتمثلة بمتوسط مربعات الخطأ (MSE) و متوسط مربعات للمقدرات (MSEθ) و معامل التحديد (R^2) و لكافة تجارب المحاكاة و يتضح ان طريقة white الحصينة اثبت انها افضل من الطريقة التقليدية حيث اثبت من

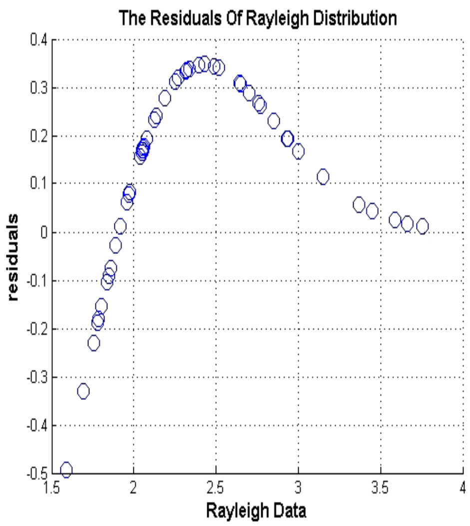
خلال المقاييس المعتمدة في المقارنة و لكافة حجوم العينات التي اعتمدت في تجارب المحاكاة .

كذلك تم رسم سلسلة البواقي (residuals) من خلال استخدام الايعاز (scatter) داخل برنامج الماتلاب .

الشكل (1) يوضح البواقي لانموذج رايلي لطريقة وايت التقليدية



الشكل (2) يوضح البواقي لانموذج رايلي لطريقة وايت الحصينة



بعد رسم البواقي لطريقتي وايت التقليدية والحصينة من اجل معرفة ايهما افضل ولنفس حجم البيانات (n=50) و

يتضح لنا ان طريقة وايت الحصينة هي افضل من الطريقة

N	$\hat{\alpha}$	$\hat{\beta}$	MSE	MSEθ	R^2
10	2.4015	2.6492	0.5113	3.2097e-004	0.5385
50	2.3998	2.6524	0.5022	2.2398e-006	0.5358
100	2.3997	2.6483	0.5046	2.6905e-007	0.5345

التقليدية من خلال مشاهدة الحد الاعلى و الحد الادنى لانتشار البواقي في كلا الشكلين (1) و (2) .

ثانيا :- البيانات فيها نسب تلوث

تم توليد البيانات فيها نسب تلوث من خلال استخدام المعادلة (15) و الجداول التالية توضح مقدرات معالم توزيع رايلي .

الجدول (3) : يوضح تقديرات معالم توزيع رايلي حسب طريقة التقليدية عندما $\beta = 1 , \alpha = 1.5$ ، $\beta = 8 , \alpha = 10$ مع تلويث

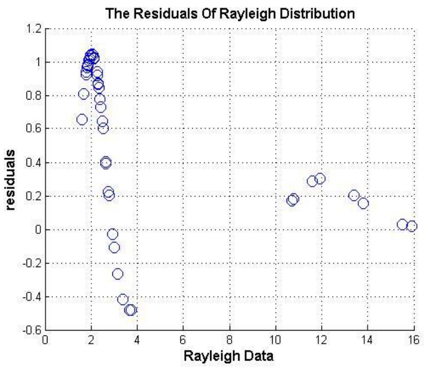
الجدول (4) : يوضح تقديرات معالم توزيع رايلي حسب طريقة الحصينية عندما $\beta = 1 , \alpha = 1.5$ ، $\beta = 8 , \alpha = 10$

N	$\hat{\alpha}$	$\hat{\beta}$	MSE	MSEθ	R^2
10	2.4013	2.6489	0.5112	3.2503e-004	0.9148
50	2.3997	2.6524	0.5022	2.2850e-006	0.9828
100	2.3997	2.6482	0.5046	2.7257e-007	0.9914

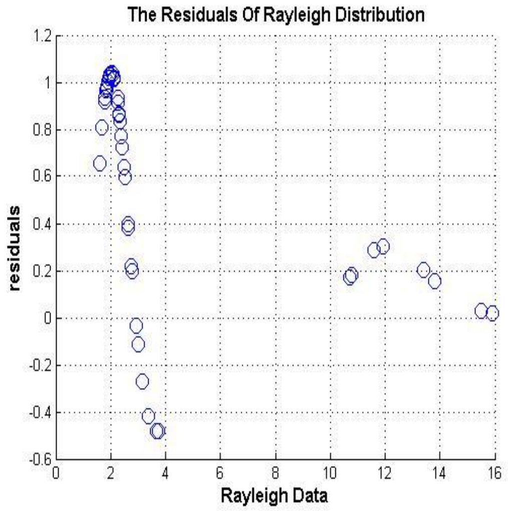
مع تلويث

من خلال متابعة النتائج المحاكاة في الجدولين (3),(4) يتضح تقارب نتائج في الطريقتين من خلال مقارنتهم معا باستخدام متوسط مربعات للخطأ (MSE)، اما مقياس متوسط مربعات للمقدرات (MSEθ) انه يبين ان طريقة وايت التقليدية افضل اما مقياس معامل التحديد (R^2) و لكافة تجارب المحاكاة و يتضح ان طريقة white الحصينة اثبت انها افضل من الطريقة التقليدية و لكافة حجوم العينات التي اعتمدت في تجارب المحاكاة .

الشكل (3) يوضح البواقي لانموذج رايلي لطريقة وايت التقليدية بوجود التلويث



الشكل (4) يوضح البواقي لانموذج رايلي لطريقة وايت الحصينة بوجود التلويث



من خلال متابعة الرسم البواقي لطريقتي وايت التقليدية و الحصينة للبيانات الملوثة في الاشكال (3) و (4) نلاحظ وجود تقارب بين الطريقتين من خلال مشاهدة الحد الاعلى و الحد الادنى لانتشارو تجمع النقاط للبواقي في كلا الاشكال (3),(4) .

الاستنتاجات :

استنادا الى نتائج اسلوب المحاكاة في الجداول (1) , … , (4) يمكن التوصل الى الاستنتاجات الاتية :-

1- من خلال متابعة نتائج المحاكاة في الجداول السابقة يتضح لنا ان طريقة وايت الحصينة هي افضل طريقة لأنها حققت اقل متوسط مربعات الخطأ (MSE) و متوسط مربعات للمقدرات (MSEθ) و معامل التحديد (R^2) لجميع الحالات و احجام العينات المستخدمة في الدراسة .

2- أظهرت نتائج المحاكاة في الجداول (3),(4) تقارب في النتائج لطريقتين بالنسبة الى متوسط مربعات الخطأ (MSE) و كما مبين في الرسم في الاشكال (3),(4) .

التوصيات :

1- نوصي باستخدام طريقة وايت الحصينة كونها تعطي نتائج جيدة .

2- نوصي بتطبيق طريقة وايت الحصينة على بيانات مخططة .

المصادر

1. الكيالي، ذياب حسين نايل (1998)، "إستخدام طريقة بيز الخبرية في تقدير معلمة النجاح في توزيع ثنائي الحدين"، إطروحة دكتوراه- كلية الإدارة والاقتصاد- الجامعة المستنصرية.

2.Afify, E.E. (2004), "Comparison of Estimators of Parameters for the Rayleigh Distribution", Faculty of Eng. Shibeen El Kom Menoufia Univ.

http://interstat.statjournals.net/YEAR/2003/articles/0302004.pdf.

3. Boudt K. and others , (2011), " Robust Explicit Estimators of Weibull Parameters " , Metrika ,Springer ,73:187-209.

4. Banks.J and others ,(2001), "Discrete-Event System Simulation" Prentice Hall. p. 3. ISBN 0-13-088702-1 .
http://en.wikipedia.org/wiki/simulation

5. Charles, E. E. (1997). "An Introduction to Reliability and Maintainability Engineering" the Mcgrau Hial, companies, Inc. New York.

6. DuMouchel, W. H., and F. L. O'Brien.(1989), "Integrating a Robust Option into a Multiple Regression Computing Environment." Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface. Alexandria, VA: American Statistical Association.

7. Davis Timothy A., Sigmon K. ,(2005) ,"MATLAB Primer ", Chapman & Hall / CRC

8. Palov Ko , A.M , (1968)."Fundamental Of Reliability Theory ".Academic Presc ,New York.

9. Street, J. O., R. J. Carroll, and D. Ruppert,(1988). "A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares." The American Statistician. Vol. 42, 152–1544.