المجلة العراقية للعلوم الإحصائية (24) 2013
ص ص[51-69]

تحديد حجم العينة البيزي لتقدير معلمة الموقع للتوزيع الطبيعي باستخدام دوال الخسارة

د. ريا سالم محمد*   واثق ناظم دهام**

المستخلص:

تم في هذا البحث تحديد حجم عينة بيز الأمثل اللازم لتقدير معلمة الموقع للتوزيع الطبيعي عندما تكون معلمة الشكل للتوزيع غير معلومة تحت دالة الخسارة التربيعية ودالة الخسارة الآسية الخطية ($LIENX$). وتم في الجانب التجريبي توضيح المقدر الأكفأ تحت دالتي الخسارة وتوضيح العلاقة بين كلفة المعاينة لكل وحدة وحجم العينة الذي تم تحديدهُ باستخدام لغة ماتلاب(Matlab).

Bayesian sample size determination for estimation of location parameter of normal distribution using loss functions

Abstract:

In this paper we determine the optimal Bayesian sample size to estimate location parameter of normal distribution when the shape parameter is unknown under squared loss function and linear exponential loss function (LIENX) .In application we explain the efficient estimator under two loss functions and obtain the relationship between the sampling cost per unit and the sample size which we determine using Matlab language.

* استاذ مساعد / قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل

** طالب ماجستير / قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل

تاريخ استلام البحث 2013/1/14 ———— تأريخ القبول 2013/3/14

1- المقدمة :

ان حجم العينة من المسائل المهمة فهناك عدة طرق لأخذ العينات من المجتمع لاستخدامها في الاستدلال الإحصائي أي تدرس خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، على أن تحمل العينة أقصى درجة من الدقة لتمثيل المجتمع المسحوبة منه. وقد توالت البحوث في هذا الموضوع لما لهُ من أهمية اذ ناقش (2006) $Sahu\ and\ Smith$[6] تحديد حجم العينة للتطبيقات في التجارب الطبية وفي التدقيق المالي تحت الامكانات الطبيعية ($normal\ likelihoods$) وقام(1997) $weiss$[9] بحساب حجم العينة للتوزيع الطبيعي بتباين معلوم باستخدام عامل بيز ($Bayes\ factor$) كما قام (1997)[2] Joseph and Bélisle بتحديد حجم العينة البيزي باستخدام عدة معايير مستندة الى الأطوال وتغطية الفترات الموثوقة اللاحقة ($posterior\ credible\ intervals$) وذلك لتقدير متوسط التوزيع الطبيعي والفرق بين متوسطي التوزيع الطبيعي وهناك من استخدم دوال الخسارة لحساب حجم العينة حيث قام $Saiful\ Islam$ (2011) [8] بتحديد حجم العينة البيزي لتقدير معلمة الموقع للتوزيع الطبيعي عندما يكون التباين للتوزيع الطبيعي معلوما تحت دوال خسارة مختلفة وبما إن أكثر الظواهر التي تسلك التوزيع الطبيعي يكون تباينها غير معلوم لذلك تم في هذا البحث تحديد حجم العينة الأمثل لتقدير معلمة الموقع للتوزيع الطبيعي عندما تكون معلمة الشكل(التباين) غير معلومة تحت دالتي الخسارة التربيعية و $LINEX$. تضمن البحث سبع فقرات في الفقرة الثانية تم توضيح نوعين من دوال الخسارة وهما دالة الخسارة التربيعية ودالة الخسارة $LINEX$ وفي الفقرة الثالثة تم توضيح دالة الكلفة وفي الفقرة الرابعة تم تحديد حجم العينة الأمثل تحت دالتي الخسارة التربيعية ودالة الخسارة $LINEX$ كما تضمنت الفقرة الخامسة حالتي المعاينة وعدم المعاينة وفي الفقرة السادسة تناولنا الجانب التجريبي حيث تم تحديد حجم العينة الامثل والمقارنة بين مقدرات معلمة الموقع تحت دالتي الخسارة التربيعية و $LINEX$ حيث تبين إن التقدير تحت دالة الخسارة التربيعية يكون أكفأ من التقدير تحت دالة الخسارة $LINEX$ وتم توضيح العلاقة بين حجم العينة المحدد وكلفة المعاينة لكل وحدة وفي الفقرة السابعة تم عرض الاستنتاجات.

2 - أنواع دوال الخسارة :

دالة الخسارة تمثل مقدار الخسارة التي نقع بها عندما نقدر المعلمة θ وان d يمثل مقدر θ.
حيث إن هناك فروقات ملحوظة بين المعلمةو القيمة التقديرية لها . والخسارة تكون دالة للفرق $(\theta - d)$

أو النسبة (d/θ) . ولمعلمة واحدة إذا ثبتنا θ يمكن إن نجد قيما مختلفة لـ d كمقدر لـ θ فإذا كانت$(d=\theta)$ فانه لا يكون لدينا خسارة وأما إذا كانت $\theta > d$ فإننا نقع بخطأ موجب ويسمى بأعلى تقدير وإذا كانت $\theta < d$ نقع بخطأ سالب ويسمى بأدنى تقدير. لذلك دالة الخسارة تعرف كدالة لـ d و θ ونرمز لها بـ $L(d,\theta)$ وهي دالة حقيقية القيمة تحقق؛ (2011) $Saiful\ Islam$[8] :

1- $0 \leq L(d,\theta)$ لكل التقديرات الممكنة لـ d للمعلمة θ تحت اختيارات عامة .

2- $L(d,\theta) = 0$ عندما $d=\theta$.

وهنالك دوال خسارة متماثلة وغير متماثلة ومن أكثر دوال الخسارة المتماثلة شيوعا هي دالة الخسارة التربيعية ومن دوال الخسارة غير المتماثلة هي دالة خسارة $LINEX$ وسوف يتم توضيحهما أدناه.

1-2- دالة الخسارة التربيعية :

وهي من دوال الخسارة المتماثلة وتعني إن كمية الخسارة المخصصة للخطأ الموجب مساوية للخطأ السالب لنفس المقدار. وهي من أهم دوال الخسارة في تقديرات بيز الإحصائية وذلك لان معظم الباحثين يميلون الى استخدام هذه الدالة نظراً لسهولة حسابها وتكون بالصيغة الآتية ; [1] $Dey\ and\ Maiti$(2012):

$$L(d,\theta) = (d - \theta)^2 \quad \cdots (1)$$

حيث إن d هو مقدر θ .

ويكون مقدر بيز تحت هذه الدالة هو متوسط التوزيع اللاحق للمعلمة المجهولة θ كالاتي; Pandey and Rao [5] (2009):

$$E\big(L(d,\theta)\big) = E(d - \theta)^2$$

$$= d^2 - 2d\, E(\theta|\underline{x}) + E(\theta^2|\underline{x})$$

وبأخذ المشتقة الأولى للمعادلة الأخيرة بالنسبة إلى d ومساواتها بالصفر نحصل على

$$\hat{d}_{sq} = E(\theta|\underline{x}) \quad \cdots (2)$$

حيث $\hat{d}_{sq}$ مقدر للمعلمة θ باستخدام دالة الخسارة التربيعية $(squared\ loss\ function)$.

2-2- دالة الخسارة $LINEX$:

هي من دوال الخسارة غير المتماثلة حيث إن استخدام دوال الخسارة المتماثلة مبني على افتراض إن الخسارة نفسها في إي اتجاه الموجب أو السالب مع ذلك فان هذا الافتراض يمكن إن لا يتحقق في العديد من الحالات العملية واستخدام دوال الخسارة المتماثلة يكون غير ملائم وفي هذه الحالات ربما يكون الخطأ الموجب أكثر أهمية من الخطأ السالب المعطى لنفس المقدار أو العكس بالعكس. لذلك سوف نستخدم دالة الخسارة LINEX وتكون بالصيغة الآتية [7] : (2006)*Soliman and Abd Ellah and Sultan*

$$L(d,\theta) = a\left[e^{b(d-\theta)} - b(d-\theta) - 1\right] \qquad \cdots(3)$$

حيث $a > 0$ هي معلمة القياس $b \neq 0$ هي معلمة الشكل وان d هو مقدر θ .

ويكون مقدر بيز للمعلمة المجهولة θ تحت دالة خسارة LINEX كالاتي ;(1986) *Zellner*[10] :

$$E\big(L(d,\theta)\big) = a\, E\left[e^{b(d-\theta)} - b(d-\theta) - 1\right]$$

وبأخذ المشتقة الأولى للمعادلة الأخيرة بالنسبة إلى d ومساواتها بالصفر وبأجراء بعض التبسيطات نحصل على مقدر بيز باستخدام دالة $LINEX$ كالأتي:

$$\hat{d}_{\text{Lin}} = -\frac{1}{b}\ \ln E(e^{-b\theta} \mid \underline{x}) \qquad \cdots(4)$$

حيث أن $\hat{d}_{\text{Lin}}$ مقدر للمعلمة θ باستخدام دالة الخسارة $Linex$.

3- دالة الكلفة:

تعتبر الكلفة من المسائل المهمة في إية دراسة والتي نهدف إلى جعلها اقل ما يمكن لذلك سوف نستخدم في هذا البحث دالة الكلفة الخطية وصيغتها هي ;(1972) *Lindley*[3] :

$$C(n) = c_\circ + cn$$

حيث إن $n > 0$

وان $C(0) = 0$

وان $c_{\circ}$ هي كلفة إعداد المعاينة أو إي كلف أخرى ذو ات علاقة بأخذ العينة .

و ان c هي كلفة المعاينة لكل وحدة.

وباستخدام أسلوب بيز لتقدير حجم العينة نحتاج إلى الكلفة الكلية $TC(n)$ والتي يمكن حسابها بالصيغة أدناه إذا كانت لا تعتمد على المشاهدات $\underline{x}$; *Saiful Islam* (2011) [8]

$$TC(n) = C(n) + PR$$

حيث إن PR تمثل دالة المخاطرة اللاحقة (*Posterior Risk*) والتي تمثل القيمة المتوقعة لدالة الخسارة. أما إذا اعتمدت دالة الكلفة الكلية على المشاهدات $\underline{x}$ ففي هذه الحالة نأخذ التوقع لدالة المخاطرة اللاحقة ونظيف إليها دالة الكلفة لنحصل على متوسط الكلفة الكلية وكالاتي:

$$E[TC(n)] = C(n) + E(PR)$$

وعليه نشتق القيمة المتوقعة للكلفة الكلية $E[TC(n)]$ بالنسبة إلى n ومساواتها بالصفر لنجد حجم عينة بيز.

4- تحديد حجم عينة بيز الأمثل:

نفرض $(x1,x2 \ldots\ldots xn)$ عينة عشوائية بحجم (n) مسحوبة من مجتمع تتوزع مفرداته بحسب دالة الكثافة الاحتمالية للتوزيع الطبيعي بمعلمة الموقع ɵ ومعلمة الشكل σ^2 . وعليه يمكن تحديد حجم عينة بيز الأمثل باستخدام دالة الخسارة التربيعية ودالة خسارة *LINEX* كالأتي:

4-1- تحديد حجم عينة بيز تحت دالة الخسارة التربيعية :

دالة الإمكان للتوزيع الطبيعي هي:

$$P(\underline{x}|\theta,\sigma^2)\ \alpha\ (\sigma^2)^{-\frac{n}{2}}\ e^{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}}$$

ودالـة الكثافـة الاحتماليـة السـابقة المرافقـة $(conjugate\ prior\ pdf)$ للمعلمتـين θ,σ^2 علـى التوالي هي:

$$\theta|\sigma^2 \sim N(\theta_\circ,\sigma^2\sigma_\circ^2)$$

حيث $\theta_\circ$ هي قيمة ابتدائية لمعلمة الموقع θ.

$\sigma_\circ^2$ هي كمية ثابتة تضرب بالتباين لتسهيل العمليات الحسابية.

$$\sigma^2 \sim \mathrm{IG}\left(\frac{a_\circ}{2},\frac{b_\circ}{2}\right)$$

حيث $b_\circ, a_\circ$ هي قيم ابتدائية لمعلمات توزيع IG لمعلمة الشكل σ^2 .

وعليه فان التوزيع اللاحق المشترك لـ σ^2 و θ يكون:

$$P(\sigma^2,\theta|\underline{x}) = \frac{\sqrt{h}}{\sqrt{2\pi}}(\sigma^2)^{-\frac{1}{2}}\ e^{-\frac{h(\theta-m)^2}{2\sigma^2}}\ \frac{\left(\frac{S^2{}_t}{2}\right)^{\frac{a_\circ+n}{2}}}{\Gamma\left(\frac{a_\circ+n}{2}\right)}\ (\sigma^2)^{-\left(\frac{a_\circ+n}{2}+1\right)}\ e^{-\frac{S^2{}_t}{2\sigma^2}}$$

حيث إن

$$S^2{}_t = b_\circ + \sum_{i=1}^{n}(x_i-\overline{x})^2 + \frac{n(\theta_\circ-\overline{x})^2}{1+n\sigma_\circ^2}$$

$$\mathrm{h} = \frac{1}{\sigma_\circ^2}+n$$

$$m = \mathrm{E}(\theta|\mathrm{x}) = \frac{\theta_\circ + n\sigma_\circ^2\overline{x}}{1+n\sigma_\circ^2} \qquad \cdots(5)$$

وان المعادلة (5) تمثل المقدر البيزي تحت دالة الخسارة التربيعية.

ونلاحظ إن

$$(\sigma^2,\theta|\underline{x}) \sim Normal\left(\mathrm{m},\frac{\sigma^2}{\mathrm{h}}\right) - Inverse\ Gamma\left(\frac{\mathrm{a_0}+\mathrm{n}}{2},\frac{S^2{}_t}{2}\right)$$

ولإيجاد التوزيع الحدي اللاحق $P(\theta|\underline{x})$ يتم كالأتي:

$$P(\theta|\underline{x}) = \int_0^\infty P(\sigma^2,\theta|\underline{x})\ d\sigma^2$$

$$P(\theta|\underline{x}) = \frac{\Gamma\left(\frac{a_\circ+n+1}{2}\right)\ \ \sqrt{h}}{\Gamma\left(\frac{a_\circ+n}{2}\right)\ \ \sqrt{\pi S_t^2}}\quad \left(1+\frac{h(\theta-m)^2}{S_t^2}\right)^{-\frac{(a_\circ+n+1)}{2}}$$

وعليه فان دالة المخاطرة اللاحقة تحت دالة الخسارة التربيعية تكون:

$$PR = \mathrm{Var}(\theta|\underline{x}) = \frac{S^2{}_t}{h(a_\circ+n-2)} \qquad \cdots(6)$$

وعليه يمكن حساب دالة الكلفة الكلية بالصيغة الآتية:

$$TC(n) = c_\circ + cn + \frac{S^2{}_t}{h(a_\circ+n-2)}$$

نلاحظ إن دالة الكلفة الكلية$TC(n)$ تعتمد على x لذلك نأخذ التوقع لـ $TC(n)$ وذلك للتخلص من x والمعلمات المجهولة (لنحصل على دالة كلفة بدلالة n فقط) فيكون متوسط الكلفة الكلية كالأتي:

$$E[TC(n)] = c_\circ + cn + \frac{b_\circ + E\left(\sum_{i=1}^n (x_i-\overline{x})^2|\,n\right) + \frac{nE((\theta_\circ-\overline{x})^2|\,n)}{1+n\sigma_\circ{}^2}}{h(a_\circ+n-2)} \qquad \cdots(7)$$

المعادلة (7) يتم إيجادها كالأتي :

كما هو معلوم أن

$$\sum_{i=1}^{n}(x_i-\overline{x})^2 \sim \sigma^2\chi^2_{(n-1)}$$

وعليه فان

$$E\left(\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\middle|\ n\right)\right) = E_{\sigma^2}\left(E_x\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\middle|\ n,\sigma^2\right)\right)$$

$$= \frac{b_\circ(n-1)}{(a_0-2)}$$

وبما ان
$$\frac{\sqrt{n}(\bar{x}-\theta_\circ)}{\sigma} \sim Normal\left(\frac{\sqrt{n}(\theta-\theta_\circ)}{\sigma}, 1\right)$$

لذلك فان
$$(\bar{x}-\theta_\circ)^2 \sim \frac{\sigma^2}{n}\, noncentral\, \chi^2_{(1)}$$

وان المعلمة اللامركزية هي $\frac{n(\theta-\theta_\circ)^2}{\sigma^2}$; (1975) *Pai Han* [4]

وعليه فان
$$E((\bar{x}-\theta_\circ)^2 \mid n)) = E_{\sigma^2} E_\theta (E_x(\theta_\circ - \bar{x})^2 \mid n, \theta, \sigma^2)$$

$$= \left(\frac{1}{n} + \sigma_\circ^2\right)\frac{b_\circ}{(a_\circ - 2)}$$

$$\therefore E(S^2{}_t \mid n) = b_\circ + \frac{n b_\circ}{(a_\circ - 2)} \qquad \cdots(8)$$

إذن متوسط الكلفة الكلية هو:

$$E\left(TC(n)\right) = c_\circ + cn + \frac{b_\circ \sigma_\circ^2}{(a_\circ - 2)(1 + n\sigma_\circ^2)} \qquad \cdots(9)$$

وباشتقاق المعادلة الأخيرة بالنسبة إلى n ومساواة المشتقة بالصفر نحصل على حجم عينة بيز الأمثل تحت دالة الخسارة التربيعية وكالاتي:

$$n_{sq}^{*} = \sqrt{\frac{b_\circ}{c(a_\circ - 2)}} - \frac{1}{\sigma_\circ^2} \qquad \cdots(10)$$

4-2- تحديد حجم عينة بيز تحت دالة الخسارة : *LINEX*

إن دالة المخاطرة اللاحقة تحت دالة الخسارة *LINEX* هي:

$$PR = E\left(L(\hat{d}, \theta)\right)$$

وبتعويض المعادلة (4) بدالة المخاطرة اللاحقة نحصل على:

$$= ab[m - \hat{d}]$$

حيث إن $\hat{d}$ و m معرفة في المعادلتين (4) و(5) على التوالي .

بعد أيجاد RP نعوضها في دالة الكلفة الكلية بنفس الطريقة السابقة.

$$TC(n) = c_0 + cn + ab\left[m + \frac{1}{b} \ln E(e^{-b\theta} \mid \underline{x})\right] \qquad \cdots(11)$$

ولإيجاد $E(e^{-b\theta} \mid \underline{x})$ نستخدم طريقة التوزيعات المركبة (*compound distributions*) لصعوبة حسابِه بطريقة مباشرة وكالاتي:

نفرض إن $\theta | w \sim N(m, \overline{\sigma^2}\, w)$

حيث إن $\overline{\sigma^2} = \frac{s^2 t}{h(a_0+n)}$

وان $w \sim IG\left(\frac{a_0+n}{2}, \frac{a_0+n}{2}\right)$

حيث إن w هو متغير افتراضي تم استخدامهُ لتسهيل العمليات الحسابية.

وعليه فان التوزيع اللاحق لـ θ المركب يكون كالأتي:

$$P(\theta | \underline{x}) = \int_0^\infty P(\theta | w, \underline{x})\, P(w | \underline{x})\, dw$$

وعليه فان $E(e^{-b\theta} \mid \underline{x}) = E_w E_\theta \left(e^{-b\theta} \mid w, \underline{x} \right)$

$$E(e^{-b\theta} \mid \underline{x}) = e^{-mb} E_w\left(e^{\frac{b^2\overline{\sigma^2} w}{2}}\right)$$

ولإيجاد $E_w(e^{\frac{b^2\overline{\sigma^2} w}{2}})$ سوف نستخدم مفكوك سلسلة ماكلورين للمقدار داخل التوقع ومنه نحصل على:

$$E_w\left(e^{\frac{b^2\overline{\sigma^2} w}{2}}\right) = \sum_{r=1}^{\infty} \frac{\left(\frac{b^2\overline{\sigma^2}}{2}\right)^r}{r!} \frac{\Gamma\left(\frac{a_0+n}{2} - r\right)}{\Gamma\left(\frac{a_0+n}{2}\right)} \left(\frac{a_0+n}{2}\right)^r$$

ونلاحظ إن المقدر $\hat{d}$ تحت دالة الخسارة $LINEX$ يكون

$$\hat{d} = m - \frac{1}{b} \ln\left[\sum_{r=1}^{\infty} \frac{\left(\frac{b^{2}\overline{\sigma^2}}{2}\right)^r}{r!} \frac{\Gamma\left(\frac{a_0+n}{2} - r\right)}{\Gamma\left(\frac{a_0+n}{2}\right)} \left(\frac{a_0+n}{2}\right)^r\right] \quad \cdots(12)$$

والآن نعوض $\hat{d}$ في الكلفة الكلية المعرفة في المعادلة (11) لنحصل على:

$$TC(n) = c_0 + cn + a \ln\left[\sum_{r=1}^{\infty} \frac{\left(\frac{b^{2}\overline{\sigma^2}}{2}\right)^r}{r!} \frac{\Gamma\left(\frac{a_0+n}{2} - r\right)}{\Gamma\left(\frac{a_0+n}{2}\right)} \left(\frac{a_0+n}{2}\right)^r\right]$$

نلاحظ إن دالة المخاطرة اللاحقة في الكلفة الكلية تعتمد على x لذلك نأخذ التوقع لـ $TC(n)$ وذلك للتخلص من x والمعلمات المجهولة (لنحصل على دالة كلفة بدلالة n فقط) فيكون متوسط الكلفة الكلية كالآتي:

$$E(TC(n)) = c_0 + cn + a\, E[\ln [\sum_{r=1}^{\infty} \frac{\left(\frac{b^{2}\overline{\sigma^2}}{2}\right)^r}{r!} \frac{\Gamma\left(\frac{a_0+n}{2} - r\right)}{\Gamma\left(\frac{a_0+n}{2}\right)} \left(\frac{a_0+n}{2}\right)^r]]$$

وباستخدام سلسلة ماكلورين وبالتعويض عن $\overline{\sigma^2}$ يكون متوسط الكلفة الكلية كالأتي:

$$E(TC(n)) = c_0 + cn + a\, E[\ln[1 + \frac{b^2 S^2{}_t}{4h(\frac{a_0+n}{2} - 1)}]]$$

$$\ln[1 + \frac{b^2 S^2{}_t}{4h\left(\frac{a_0+n}{2} - 1\right)}] = \frac{b^2 S^2{}_t}{4h\left(\frac{a_0+n}{2} - 1\right)} - \left(\frac{b^2 S^2{}_t}{4h\left(\frac{a_0+n}{2} - 1\right)}\right)^2 + \cdots$$

$$\text{as} \left|\frac{b^2 S^2{}_t}{4h(\frac{a_0+n}{2} - 1)}\right| < 1$$

وبأخذ الحد الأول من سلسلة ماكلورين كتقريب ولتسهيل الحل نحصل على

$$E(TC(n)) = c_o + cn + \frac{ab^2 b_o \sigma_o^2}{2(a_o - 2)(1 + n\sigma_o^2)} \qquad \cdots(13)$$

وباشتقاق متوسط الكلفة الكلية بالنسبة إلى n ومساواة المشتقة بالصفر نحصل على حجم عينة بيز الأمثل تحت دالة خسارة $LINEX$ وكالاتي:

$$n_{\text{lin}}^* = \sqrt{\frac{ab^2 b_o}{2c(a_o - 2)}} - \frac{1}{\sigma_o^2} \qquad \cdots(14)$$

5- حالة المعاينةو عدمها; $Saiful\ Islam$[8] (2011) :

إن أسلوب بيز في التقدير يعتمد على حساب الاحتمال اللاحق للمعلمة ولإيجاد هذا الاحتمال نحتاج إلى معلومات من العينة ومن الاحتمال السابق حيث إن:

المعلومات اللاحقة = المعلومات السابقة + معلومات العينة

المعلومات اللاحقة يتم إيجادها في حالتين هما:

1- إذا كان الاحتمال السابق هو السائد $(dominant)$ فان اغلب المعلومات اللاحقة تأتي من الاحتمال السابق وعليه في مثل هذه الحالة قد لا نحتاج إلى معلومات المشاهدات وهذه الحالة تسمى حالة عدم المعاينة $(no\ sampling)$.

2- إذا كانت دالة الإمكان هي السائدة $(dominant)$ فسوف نحتاج إلى معلومات المشاهدات لذلك نقوم بأخذ عينة وتسمى حالة المعاينة$(sampling)$.

ولنبين حالة الاعتماد على المعاينة أو عدمه لتقدير معلمة الموقع للتوزيع الطبيعي تحت دالة الخسارة التربيعية ودالة خسارة $LINEX$ وذلك بالاعتماد على كلفة المعاينة لكل وحدة.

1-5- تحت دالة الخسارة التربيعية :

في حالة عدم اخذ عينة $(n^* = 0)$ فان متوسط الكلفة الكلية تحت دالة الخسارة التربيعية الموضح في المعادلة (9) يكون:

$$E(TC(0)) = \frac{b_o \sigma_o^2}{(a_o - 2)}$$

وعند اخذ عينة $(n^* = n^*_{sq})$ فان متوسط الكلفة الكلية يكون:

$$E(TC(n)) = c_o - \frac{c}{\sigma_o^2} + 2\sqrt{\frac{cb_o}{(a_o - 2)}}$$

لذلك فان حجم عينة بيز الأمثل بصورة عامة:

$$n^* = \max\left[0 \ , \sqrt{\frac{b_o}{c(a_o - 2)}} - \frac{1}{\sigma_o^2}\right]$$

وعليه فإننا نقرر اخذ عينة (حالة المعاينة)عندما يكون المقدار $\left(\sqrt{\frac{b_o}{c(a_o-2)}} - \frac{1}{\sigma_o^2}\right)$ موجبا ويكون

متوسط الكلفة الكلية في حالة المعاينة اقل من متوسط الكلفة الكلية في حالة عدم المعاينة:

$$\left[\frac{b_o \sigma_o^2}{(a_o - 2)} > c_o - \frac{c}{\sigma_o^2} + 2\sqrt{\frac{cb_o}{(a_o - 2)}} \ and \ \sqrt{\frac{b_o}{c(a_o - 2)}} > \frac{1}{\sigma_o^2}\right] \quad \cdots (15)$$

ونقرر عدم اخذ عينة (حالة عدم المعاينة) عندما يكون المقدار $\left(\sqrt{\frac{b_o}{c(a_o-2)}} - \frac{1}{\sigma_o^2}\right)$ سالب

أو عندما يكون متوسط الكلفة الكلية في حالة المعاينة اكبر من متوسط الكلفة الكلية في حالة عدم المعاينة:

$$\left[\frac{b_o \sigma_o^2}{(a_o - 2)} < c_o - \frac{c}{\sigma_o^2} + 2\sqrt{\frac{cb_o}{(a_o - 2)}} \ and \ \sqrt{\frac{b_o}{c(a_o - 2)}} > \frac{1}{\sigma_o^2}\right]$$

$$or \ \sqrt{\frac{b_o}{c(a_o - 2)}} \leq \frac{1}{\sigma_o^2}$$

وبالتعويض عن $\sqrt{c}$ بـ y وبحل المتباينة (15) نحصل على :

$$\frac{y^2}{\sigma_o^2} - 2y\sqrt{\frac{b_o}{(a_o-2)}} + \frac{b_o\sigma_o^2}{(a_o-2)} > c_o \; and \; y < \sqrt{\frac{b_o\sigma_o^4}{(a_o-2)}}$$

وعليه فأننا نقوم بالمعاينة عندما تكون قيمة c محددة بالفترة الآتية

$$c < [\sqrt{\frac{b_o\sigma_o^4}{(a_o-2)}} - \sqrt{c_o\sigma_o^2}\;]^2 \quad]$$

حيث إن($a_o, b_o, c_o, \sigma_o^2$) تكون معلومة فإذا كانت كلفة المعاينة لكل وحدة تتجاوز المقدار $[\sqrt{\frac{b_o\sigma_o^4}{(a_o-2)}} - \sqrt{c_o\sigma_o^2}\;]^2$ فان المعاينة ليست ذات قيمة بسبب ارتفاع كلفة المعاينة لكل وحدة وفي هذه الحالة يتم الاعتماد فقط على المعلومات السابقة لتقدير معلمة الموقع للتوزيع الطبيعي عندما تكون معلمة الشكل للتوزيع الطبيعي غير معلومة.

2-5-تحت دالة خسارة $LINEX$:

في حالة عدم اخذ عينة $(n^* = 0)$ فان متوسط الكلفة الكلية تحت دالة خسارة $LINEX$ الموضح في المعادلة (13) يكون:

$$E(TC(0)) = \frac{ab^2 b_o\sigma_o^2}{2(a_o-2)}$$

وعند اخذ عينة $(n^* = n^*_{lin})$ فان متوسط الكلفة الكلية يكون:

$$E(TC(n)) = c_o - \frac{c}{\sigma_o^2} + 2\sqrt{\frac{cab^2b_o}{2(a_o-2)}}$$

لذلك فان حجم عينة بيز الأمثل بصورة عامة :

$$n^* = \max\left[0 \;,\; \sqrt{\frac{ab^2b_o}{2c(a_o-2)}} - \frac{1}{\sigma_o^2}\right]$$

وعليه فإننا نقرر اخذ عينة (حالة المعاينة) في الحالة الآتية

$$\left[\frac{ab^2 b_o \sigma_o^2}{2(a_o-2)} > c_0 - \frac{c}{\sigma_0^2} + 2\sqrt{\frac{cab^2 b_o}{2(a_o-2)}} \; and \; \sqrt{\frac{ab^2 b_o}{2c(a_o-2)}} > \frac{1}{\sigma_o^2}\right] \qquad \cdots(16)$$

ونقرر عدم اخذ عينة (حالة عدم المعاينة) في الحالة الآتية

$$\left[\frac{ab^2 b_o \sigma_o^2}{2(a_o-2)} < c_0 - \frac{c}{\sigma_0^2} + 2\sqrt{\frac{cab^2 b_o}{2(a_o-2)}} \; and \; \sqrt{\frac{ab^2 b_o}{2c(a_o-2)}} > \frac{1}{\sigma_o^2}\right]$$

$$or \; \sqrt{\frac{ab^2 b_o}{2c(a_o-2)}} \le \frac{1}{\sigma_o^2}$$

وبالتعويض عن $\sqrt{c}$ ب y وبحل المتباينة (16) وبنفس الأسلوب السابق فنقوم بالمعاينة عندما تكون قيمة c محددة بالفترة الآتية

$$c \; < \left[\sqrt{\frac{ab^2 b_o \sigma_0^4}{2(a_o-2)}} - \sqrt{c_0 \sigma_o^2}\;\right]^2$$

حيث إن $(a_0, b_0, c_0, \sigma_0{}^2, a, b)$ تكون معلومة فإذا كانت كلفة المعاينة لكل وحدة تتجاوز المقدار $\left[\sqrt{\frac{ab^2 b_o \sigma_0^4}{2(a_o-2)}} - \sqrt{c_0 \sigma_o^2}\;\right]^2$ فان المعاينة ليست ذات قيمة بسبب ارتفاع كلفة المعاينة لكل وحدة وفي هذه الحالة يتم الاعتماد فقط على المعلومات السابقة لتقدير معلمة الموقع للتوزيع الطبيعي عندما تكون معلمة الشكل للتوزيع الطبيعي غير معلومة.

6- الجانب التجريبي:

في هذا البحث تم توليد إعداد عشوائية تتبع التوزيع الطبيعي باستخدام لغة ماتلاب $(Matlab)$ وبتنفيذ برنامج لحساب المقدرات المبينة في المعادلتين(5)و (12) ثم نقارن بين مقدرات معلمة الموقع للتوزيع الطبيعي تحت دالتي الخسارة التربيعية و $LINEX$ عن طريق معيار MSE :

$$mse(\hat{\theta}) = \frac{\sum_{i=1}^{R}(\hat{\theta}_i - \theta)^2}{R} \qquad \cdots(17)$$

حيث إن عدد المكررات $R = 500$.

وسوف تتم هذه المقارنة في حالتين في الحالة الأولى سوف يكون حجم العينة الذي تم استخدامهُ لحساب معلمة الموقع تحت دالتي الخسارة هو حجم العينة البيزي الأمثل المحسوب في المعادلتين (10) و(14) وتم حساب قيم معيار MSE والكفاءة النسبية للمقدرات وكما موضح في الجدول (1) و(2) وفي الحالة الثانية تم استخدام أحجام عينات متساويا لحساب مقدرات معلمة الموقع تحت دالتي الخسارة وأحجام العينات المستخدمة في هذه الحالة هي$(20,50,100)$ وتم حساب قيم معيار MSE والكفاءة النسبية للمقدرات وكما موضح في الجدول(3) وتم استخدام نفس قيم المعلمات والقيم الابتدائية الموضحة أدناه للحالتين وقيم c تستخدم فقط في حساب عينة بيز للحالة الأولى.

وتم اخذ قيم لـ θ هي 1.5 ,0.5 واستخدامها في المعادلة (17) وقيم لـ a هي0.6, 0.5, 0.4 واستخدامها في المعادلة (14) وقيم لـ b هي 2, 1, -1,-2 واستخدامها في المعادلتين(12) و(14) وقيم لـ c هي 0.001 ,0.01 واستخدامها في المعادلتين (10) و(14) وقيم مختلفة من θ و a و b و c وقيم أولية لـ ($a_0 = 3$ و $b_0 = 4$ و $\sigma_0^2 = 0.4$ و $\theta_0 = 0.2$). من الجداول(1) و(2) نلاحظ إن حجم العينة تحت دالة خسارة $LINEX$ يزداد بازدياد قيمة a و القيمة المطلقة لـ b. وفي الحالتين للجداول$(1-3)$ فان قيمة MSE تحت دالة الخسارة التربيعية هي اقل من قيمة MSE تحت دالة خسارة $LINEX$ عندما تكون قيمة b موجبة وبذلك الكفاءة تكون النسبية اقل من واحد وعندما تكون قيمة b سالبة فان قيمة MSE تحت دالة خسارة $LINEX$ هي اقل من قيمة MSE تحت دالة الخسارة التربيعية وبذلك تكون الكفاءة النسبية اكبر من واحد . ونلاحظ عندما يكون حجم العينة كبير تكون الكفاءة متماثلة لدالتي الخسارة. وان حجم العينة يتناقص كلما ارتفعت كلفة المعاينة لكل وحدة بصورة عامة لدالتي الخسارة .

الجدول (1) يمثل قيما لحجم عينة بيز ومعيار MSE والكفاءة النسبية عندما $\mathbf{a = 0.4}$

θ	b	c	n^*_{sq}	n^*_{lin}	mse_{lin}	mse_{lin}	$R(\hat{d}_{sq}, \hat{d}_{lin})$
0.5	2	0.01	17	15	0.0477	0.0552	0.8641
		0.001	60	54	0.0149	0.0156	0.9545
	1	0.01	17	6	0.0488	0.0579	0.8442
		0.001	60	25	0.0147	0.0156	0.9413
	-1	0.01	17	6	0.0428	0.0421	1.0147
		0.001	60	25	0.0156	0.0156	0.9992
	-2	0.01	17	15	0.0471	0.0455	1.0348
		0.001	60	54	0.0160	0.0159	1.0094
1.5	2	0.01	17	15	0.0673	0.0930	0.7234
		0.001	60	54	0.0184	0.0209	0.8774
	1	0.01	17	6	0.0707	0.1037	0.6816
		0.001	60	25	0.0185	0.0215	0.8604
	-1	0.01	17	6	0.0744	0.0536	1.3898
		0.001	60	25	0.0180	0.0162	1.1152
	-2	0.01	17	15	0.0789	0.0593	1.3305
		0.001	60	54	0.0180	0.0162	1.1114

الجدول (2) يمثل قيما لحجم عينة بيز ومعيار MSE والكفاءة النسبية عندما $\mathbf{a = 0.6}$

θ	b	c	n^*_{sq}	n^*_{lin}	mse_{lin}	mse_{lin}	$R(\hat{d}_{sq}, \hat{d}_{lin})$
0.5	2	0.01	17	19	0.0464	0.0520	0.8928
		0.001	60	66	0.0139	0.0144	0.9662
	1	0.01	17	8	0.0461	0.0523	0.8820
		0.001	60	32	0.0152	0.0157	0.9686
	-1	0.01	17	8	0.0469	0.0460	1.0185
		0.001	60	32	0.0163	0.0157	1.0406
	-2	0.01	17	19	0.0453	0.0436	1.0376

		0.001	60	66	0.0146	0.0145	1.0072
1.5	2	0.01	17	19	0.0755	0.0984	0.7681
		0.001	60	66	0.0176	0.0197	0.8970
	1	0.01	17	8	0.0694	0.0966	0.7189
		0.001	60	32	0.0161	0.0180	0.8962
	-1	0.01	17	8	0.0735	0.0560	1.3128
		0.001	60	32	0.0190	0.0178	1.0658
	-2	0.01	17	19	0.0759	0.0608	1.2486
		0.001	60	66	0.0161	0.0148	1.0881

الجدول (3) يمثل قيم معيار MSE والكفاءة النسبية عندما $\mathbf{a = 0.5}$

n	θ	b	mse_{lin}	mse_{lin}	$R(\hat{d}_{sq}, \hat{d}_{lin})$
20	0.5	2	0.0406	0.0449	0.9040
		1	0.0358	0.0377	0.9504
		-1	0.0449	0.0438	1.0240
		-2	0.0371	0.0364	1.0186
	1.5	2	0.0621	0.0788	0.7883
		1	0.0569	0.0644	0.8841
		-1	0.0670	0.0606	1.1057
		-2	0.0642	0.0526	1.2212
50	0.5	2	0.0165	0.0173	0.9533
		1	0.0229	0.0232	0.9890
		-1	0.0168	0.0165	1.0211
		-2	0.0165	0.0165	1.0273
	1.5	2	0.0197	0.0223	0.8847
		1	0.0233	0.0247	0.9451
		-1	0.0231	0.0219	1.0573
		-2	0.0214	0.0195	1.0990

100	0.5	2	0.0092	0.0093	0.9833
		1	0.0092	0.0092	0.9840
		-1	0.0098	0.0097	1.0043
		-2	0.0094	0.0094	1.0081
	1.5	2	0.0104	0.0111	0.9399
		1	0.0093	0.0097	0.9666
		-1	0.0109	0.0106	1.0232
		-2	0.0108	0.0104	1.0374

7-الاستنتاجات:

1- تم تحديد حجم عينة بيز الأمثل لتقدير معلمة الموقع للتوزيع الطبيعي عندما تكون معلمة الشكل غير معلومة وعليه يكون اتخاذ القرار بشأن المعاينة وعدمها تحت دالة الخسارة التربيعية ودالة خسارة $LINEX$ كالاتي:

تحت دالة الخسارة التربيعية يجب اخذ عينة (وجود المعاينة) عندما تكون كلفة المعاينة لكل وحدة اقل من المقدار $[\sqrt{\frac{b_o\sigma_o^4}{(a_o-2)}}-\sqrt{c_o\sigma_o^2}\,]^2$. وتحت دالة خسارة $LINEX$ يجب اخذ عينة (وجود المعاينة) عندما تكون كلفة المعاينة لكل وحدة اقل من المقدار $[\sqrt{\frac{ab^2b_o\sigma_o^4}{2(a_o-2)}}-\sqrt{c_o\sigma_o^2}\,]^2$.

وبخلاف ما تقدم لدالتي الخسارة التربيعية و$LINEX$ فلا نقوم بأخذ عينة (عدم المعاينة).

2- اتضح إن حجم العينة المحدد يقل كلما ازدادت كلفة المعاينة لكل وحدة ويزداد هذا التناقص بزيادة b_o وانخفاض a_o باستخدام دالة الخسارة التربيعية ويزداد هذا التناقص وأيضا عندما نثبت a_o و b_o مع زيادة a أو b أو كليهما باستخدام دالة خسارة $LINEX$.

3- عند مقارنة النتائج باستخدام معيار MSE تبين إن مقدر بيز تحت دالة الخسارة التربيعية يكون أفضل من مقدر بيز تحت دالة خسارة $LINEX$ عندما تكون $(b > 0)$ وبالعكس يكون مقدر بيز تحت دالة خسارة $LINEX$ هو الافضل في استخدام عينة بيز او حجم عينة ثابت لحساب المقدرات .

4- وأظهرت النتائج تقارباً بين مقدرات بيز لدالتي الخسارة بالنسبة لحجم العينة الكبير .

المصادر:

1- Dey, S, Maiti, S. S. (2012);" Bayesian estimation of the parameter of Rayleigh distribution under the extended Jeffrey ′s prior"; Electron. J. App. Stat. Anal., Vol. 5, Issue 1, 44 – 59.

2- Joseph, L, Bélisle, P. (1997);"Bayesian sample size determination for normal means and differences between normal means ";Statistician,46, 2, 209-226.

3- Lindley, D. V. (1972);" Bayesian statistics, a review"; Philadelphia: Society for Industrial and Applied Mathematics.

4- Pai Han, c.(1975);" Some Relationships Between Noncentral Chi-Squared and Normal Distributions "; Biometrika, Vol. 62, No. (1) , pp. 213-214.

5- Pandey, H, Rao, A. K. (2009);" Bayesian estimation of the shape parameter of a generalized Pareto distribution under asymmetric loss function "; Hacettepe Journal of Mathematics and Statistics, Vol 38 (1), 69 – 83.

6- Sahu, S. K. and Smith, T. F. M. (2006); "A Bayesian method of sample size determination with practical applications". determination with practical applications. Journal of the Royal Statistical Society A, 169, Part-2, 235-253.

7- Soliman, A. A, Abd Ellah, A. H, Sultan, K. S. (2006);" Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approaches"; Computational Statistics and Data Analysis 51, 2065 – 2077.

8- Saiful Islam, A. F. M. (2011);" Loss functions, utility functions and Bayesian sample size determination "; PhD Thesis, University of London.

9- Weiss, R. (1997); "Bayesian sample size calculation for hypothesis testing". The Statistician, 46, 2, 185-191.

10- Zellner, A. (1986);" Bayesian estimation and prediction using asymmetric loss functions"; Journal of the American Statistical Associations, Vol.81, No.394, 446-451.