 ISSN : 1813 – 1662

دراسة محاكاة لنظرية الغاية المركزية للعوائل الاسية

وائل عبد اللطيف جاسم
المعهد التقني ، الموصل ، العراق

(تاريخ الاستلام: ٢٧ / ١٠ /٢٠٠٨ ، تاريخ القبول: ١٢ / ٤ / ٢٠٠٩)

الملخص :

يظهر هذا البحث انه تحت شروط الانتظام واذا كان $X_1,.......,X_n$ هي متغيرات عشوائية متماثلة التوزيع ومستقلة من k من المعلمات للعائلة الاسية المنتظمة باحصاءة تامة وكافية T_n مع $E(T_n)=\mu_T$ فان $\sqrt{n}(T_n-\mu_T)\sim N_k(0,I(\eta))$ ، اذ ان $I(\eta)$ هي مصفوفة المعلومات للمعلمة الطبيعية للعائلة الاسية. كذلك يمكن باستخدام طريقة دلتا في ايجاد التوقع والتباين للدوال المعقدة . لقد تم في هذا البحث توليد اعداد عشوائية لتوزيعين مستمرين وتوزيعين متقطعين يعودون للعائلة الاسية وقد اظهرت نتيجة المحاكاة بان نظرية الغاية المركزية تعطي تقريب جيد للتوزيع الطبيعي .

المقدمة

العائلة لدالة الكثافة الاحتمالية (p.d.f) او دالة الكتلة الاحتمالية (p.m.f) $\{f(y|\theta):\theta=(\theta_1,....,\theta_j)\in\Theta\}$ للمتغير العشوائي Y هي عائلة اسية اذا كانت

$$f(y|\theta)=h(y)c(\theta)\exp[\sum_{i=1}^{k}w_i(\theta)t_i(y)] \qquad (1)$$

لـ $y\in Y$ حيث $c(\theta)\geq 0$ و $h(y)\geq 0$. والدوال c,h,t_i,w_i هي دوال ذات قيم حقيقية . من التعريف نلاحظ ان $c,w_1,...,w_k$ لا تعتمد على y وان $h,t_1,...,t_k$ لا تعتمد على θ . الدعم للتوزيع هو y وان فضاء المعلمة هو Θ. العائلة هي عائلة اسية بـ k من المعلمات اذا كانت k هي اصغر عدد صحيح عند تطبيق المعادلة (١) .

فاذا فرضنا ان $Y_1,.......,Y_n$ هي توزيعات عشوائية مستقلة ومتماثلة ، فان التوزيع المشترك لـ $Y_1,.......,Y_n$ يتبع ايضا العائلة الاسية أي ان :

$$f(y_1,.....,y_n)=\prod_{i=1}^{n}f(y_i)=[\prod_{i=1}^{n}h(y_i)][c(\theta)]^n.\prod_{i=1}^{n}\exp[\sum_{j=1}^{k}w_j(\theta)t_j(y_i)] \qquad (2)$$

$$=[\prod_{i=1}^{n}h(y_i)][c(\theta)]^n.\exp[\sum_{j=1}^{k}w_j(\theta)(\sum_{i=1}^{n}t_j(y_i))]$$

من معادلة (2) ليكن بان $h^*(y_1,.....,y_n)=\prod_{i=1}^{n}h(y_i)$ و $c^*(\theta)=[c(\theta)]^n$ و $w_j^*(\theta)=w_j(\theta)$ و $t_j^*(y_1,.....,y_n)=\sum_{i=1}^{n}t_j(y_i)$.

ان استخدام المعلمة الطبيعية η مفيد جدا عند تطبيق المعادلة (١) ، فلو فرضنا ان Ω هو فضاء المعلمة الطبيعية لـ η ، فان المعلمة الطبيعية للعائلة الاسية ستكون كما يلي :

$$f(y|\eta)=h(y)c^*(\eta)\exp[\sum_{i=1}^{k}\eta_i t_i(y)] \qquad (3)$$

جيث ان $h(y)$ و $t_i(y)$ هي نفسها بالمعادلة (١) وان $\eta\in\Omega$.

ان العائلة الاسية بمعلمة واحدة $f(y|\theta)=h(y)c(\theta)\exp[w(\theta)t(y)]$ لها معلمة طبيعية وكالاتي :

$$f(y|\eta)=h(y)c^*(\eta)\exp[\eta t(y)]$$

سوف نستعرض الان العائلة الاسية المنتظمة والعائلة الاسية الكاملة .

افرض ان $d_i(y)$ تعرف $w_i(\theta),t_i(y)$ او η_i . قيد الخطية يتحقق بـ $d_1(x),.....,d_k(x)$ اذا كان $\sum_{i=1}^{k}a_i d_i(x)=c$ لبعض الثوابت a_i و c ولجميع قيم X في العينة او فضاء المعلمة عندما لا تكون جميع قيم $a_i=0$.

اذا كان $\sum_{i=1}^{k}a_i d_i(x)=c$ لجميع قيم x فقط واذا كان $a_1=.......=a_k=0$ فان $d_i(x)$ لا تحقق قيد الخطية (انظر Johanson (1979, p. 3)) .

في الجبر الخطي يمكن القول بان $d_i(x)$ هي مستقلة خطيا اذا لم تحقق شرط الخطية .

افرض ان $\tilde{\Omega}$ هي مجموعة محددة عندما يكون تكامل نواة الدالة محدد وكما يلي :

$$\tilde{\Omega}=\{\eta=(\eta_1,.....,\eta_k):\frac{1}{c^*(\eta)}\equiv\int_{-\infty}^{\infty}h(y)\exp[\sum_{i=1}^{k}\eta_i t_i(y)]dy<\infty\} \qquad (4)$$

ويبدل التكامل بالمجموع في حالة دالة الكتلة الاحتمالية .

الشروط مهمة جدا لمعرفة فيما اذا كانت العائلة الاسية منتظمة او كاملة وهي :

١- فضاء المعلمة الطبيعي $\Omega=\tilde{\Omega}$.

٢- η_i و t_i لا تحقق شرط الخطية للمعلمة الطبيعية .

٣- Ω هي مجموعة مفتوحة ذات بعد k .

فاذا كانت الشروط الثلاثة متحققة فان العائلة الاسية هي عائلة اسية منتظمة (REF) ، واذا تحقق الشرطان الاول والثاني فقط فان العائلة الاسية تسمى عائلة اسية كاملة .

نلاحظ من تعريف Ω وΘ ان المعادلتين (١) و (٣) تتحقق لكل $\theta\in\Theta$ ولكل $\eta\in\Omega$.

ان المعلمة الطبيعية للتوزيعات المستمرة والمتقطعة يجب ان تكون موجودة ، فعلى سبيل المثال في توزيع ثنائي الحدين $bin(n,p)$ وعلى فرض ان n معلومة ، فان المعلمة الطبيعية للتوزيع هي $\eta=\log(p/(1-p))$ ونلاحظ انه بدلا من استخدام $\Theta=[0,1]$ نستخدم $p\in\Theta=(0,1)$ وعليه فان $\eta\in\Omega=(-\infty,\infty)$.

يمكننا القول بان توزيع معين يتبع العائلة الاسية بـ k من المعلمات وذلك بوضع $\eta_i=w_i(\theta)$ ، ومن ثم نفرض ان فضاء المعلمة الطبيعية هو $\Omega=\{(\eta_1,.....,\eta_k):\eta_i=w_i(\theta),\theta\in\Theta\}$.

ان $\eta_i=w_i(\theta)$ تكون صحيحة لمعظم التوزيعات لان η هي دالة متباينة بالنسبة لـ θ وللمزيد انظر (Olive,2005) . لمعرفة فيما اذا

كانت Ω هي مجموعة مفتوحة ذات بعد k يجب برهنة ان Ω هي حاصل ضرب فترات مفتوحة وكما موضح بالمثال التالي :

مثال : افرض ان $f(y|\mu,\sigma^2)$ تتبع التوزيع الطبيعي $N(\mu,\sigma^2)$ حيث ان $\theta=(\mu,\sigma)$ وان $-\infty<\mu<\infty$ ، $\sigma>0$.

$$f(y|\theta)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

$$f(y|\mu,\sigma^2)=\underbrace{\boxed{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\mu^2}{2\sigma^2}}}}_{c(\mu,\sigma)\ge 0}.e^{(\underbrace{-\frac{1}{2\sigma^2}}_{w_1(\theta)}\overbrace{y^2}^{t_1(y)}+\underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)}\overbrace{y}^{t_2(y)})}$$

نلاحظ ان التوزيع الطبيعي هو عائلة اسية بمعلمتين $\eta_2=\frac{\mu}{\sigma^2}$ ، $\eta_1=\frac{-0.5}{\sigma^2}$ اذا كانت $\sigma>0$ وان $\Omega=(-\infty,0)\times(-\infty,\infty)$ ، لذلك فان العائلة هي عائلة اسية منتظمة بمعلمتين ، انظر (Olive,2005) .

ان Casella and Berger (2002) ، Barndorff–Nielsen (1982) ، Cox and Hinkley (1974) ، Lehmann (1983) ، Schervish (1995) واخرين كثيرين اقترحوا بانه تحت شروط الانتظام ، اذا كان $Y_1,.......,Y_n$ هي توزيعات عشوائية مستقلة ومتشابهة (لها نفس التوزيع) من عائلة اسية منتظمة بمعلمة واحدة وان $\hat{\theta}$ هي مقدر الامكان الاعظم لـ θ ، فان

$$\sqrt{n}(\phi(\hat{\theta})-\phi(\theta)) \sim N[0,FCRLB_1(\phi(\theta))] \qquad (5)$$

حيث ان فريشي كرامر – راو لـ $\phi(\theta)$ هي

$$FCRLB_1(\phi(\theta))=\frac{[\phi'(\theta)]^2}{I_1(\theta)}$$

ومعلومات فيشر التي تعتمد على حجم عينة واحدة هي

$$I_1(\theta)=-E_\theta[\frac{\partial^2}{\partial\theta^2}\log(f(y|\theta))]$$

اما بالنسبة لنظرية الغاية المركزية فقد عرفت في القرن الثامن عشر لكن التطور الدقيق لهذه النظرية يعود الى علماء الرياضيات الروسيين وبالأخص شبيشيف ، ماركوف وليابنوف وقد بلغت ذروتها في بداية القرن العشرين .

توصف نظرية الغاية المركزية بانها النظرية الاكثر اهمية في الاحصاء ، وتنص النظرية على انه اذا كانت لدينا عينات عشوائية مستقلة بحجم n من توزيع معين له وسط μ وتباين σ^2 محددين (finite) . فان متوسطات العينات تتوزع محاذيا (عندما $n\rightarrow\infty$) توزيعا طبيعيا بوسط μ وتباين σ^2/n .

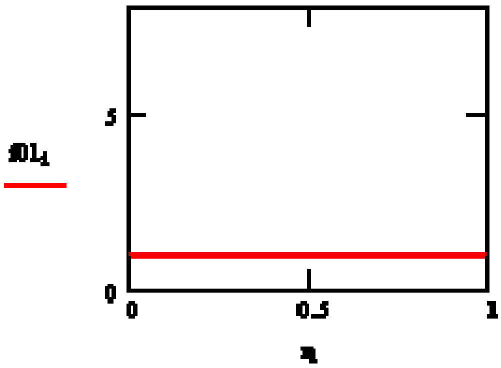
للمزيد انظر (Wackerly, Mendenhall III and Scheaffer,2002)

ان المهم في نظرية الغاية المركزية هو ان عناصر المتجهات هي متغيرات عشوائية .

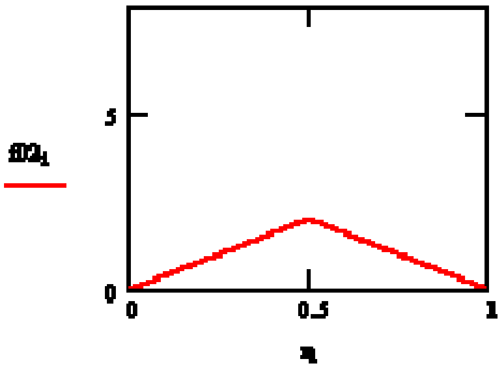
نظرية الغاية المركزية وطريقة دلتا

Central Limit Theorem & Delta Method

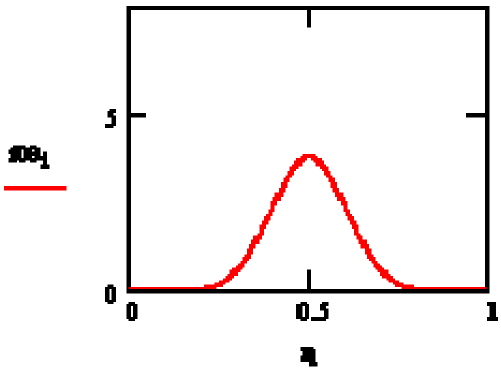
ان نظرية الغاية المركزية تعتبر واحدة من النظريات المهمة في الاحصاء الرياضي وهي تستخدم تقريبا حيثما يستخدم الاحصاء الرياضي . ان الفائدة المتوخاة من استخدام هذه النظرية تتمثل بتعريف بسيط هو ان نظرية الغاية المركزية في حالة توفر بعض الشروط المعينة فان توزيع الوسط الحسابي لعدد من المتغيرات العشوائية المستقلة يقترب من التوزيع الطبيعي عندما يقترب عدد المتغيرات العشوائية الى المالانهاية (Infinity) .



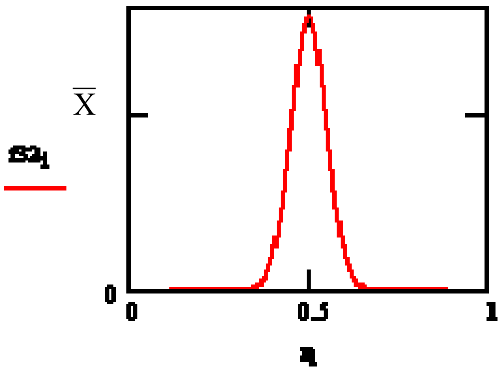
NonNormal Distribution of X



Distribution of Xbar when sample size is 2



Distribution of Xbar when sample size is 8



Distribution of Xbar when sample size is 32

ان جمال نظرية الغاية المركزية تكمن في بساطتها ولتوضيح ذلك نأخذ عينات من التوزيع المنتظم (Uniform Distribution) وبحجم عينة ١ ، ٢ ، ٨ ، ٣٢ وكما موضح بالشكل رقم (١) : (انظر Le Cam, L. (1986))

ان من فروض تطبيق نظرية الغاية المركزية هي :

١– ان تكون العينة المسحوبة من المجتمع لها توزيع عشوائي مستقل ومتماثل.

٢– المتغير المسحوب من العينة يجب ان يمتلك وسط وتباين مجتمع محدد (finite) .

فاذا ما توفرت الفروض اعلاه فان :

١– متوسط العينة يتوزع توزيعا طبيعيا مهما يكن توزيع المتغير الأصلي $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$.

٢– متوسط العينة له نفس القيمة المتوقعة لمتوسط المجتمع μ أي ان $E(\bar{X})=\mu$.

٣– الانحراف المعياري لمتوسط العينة هو σ^2/n أي ان $V(\bar{X})=\frac{\sigma^2}{n}$

نظرية (١) : طريقة دلتا Delta Method

اذا كان $g'(\theta)\neq 0$ وان $\sqrt{n}(T_n-\theta)\sim N(0,\sigma^2)$ فان

$$\sqrt{n}(g(T_n)-g(\theta))\sim N(0,\sigma^2[g'(\theta)]^2)$$

الاحصاءة الكافية والعائلة الاسية Sufficient Statistics & Exponential Family

الفكرة الاساسية للاحصاءة الكافية T(y) لـ θ هو ان كل المعلومات المطلوبة من البيانات $Y_1,\ldots\ldots,Y_n$ حول المعلمة θ هي موجودة في الاحصاءة T(y) . الاحصاءة الكافية T(y) هي اصغر احصاءة كافية اذا كان لاي احصاءة كافية اخرى T'(y) ، T(y) هي دالة لـ T'(y) .

سوف نتكلم الان عن نظريتين مهمتين للاحصاءة الكافية :

أ– **النظرية العاملية : Factorization Theorem**

افرض ان $f(x|\theta)$ دالة الكثافة الاحتمالية او دالة الكتلة الاحتمالية للعينة x . الاحصاءة T(x) هي احصاءة كافية لـ θ اذا كان :

$$f(x|\theta)=g(\mathrm{T}(x)|\theta)h(x)$$

اذ ان g و h هي دوال غير سالبة ، انظر (Olive,2006) .

البرهان :

اذا كان لدينا $X_1,\ldots\ldots,X_n$ هي توزيعات عشوائية متماثلة (identical) من العائلة الاسية فان

$$f(x|\theta)=h(x)c(\theta)\exp[w_1(\theta)t_1(x)+\ldots\ldots+w_k(\theta)t_k(x)]$$

وباستخدام المعلمة الطبيعية فان

$$f(x|\eta)=h(x)c^*(\eta)\exp[\eta_1 t_1(x)+\ldots\ldots+\eta_k t_k(x)]$$

لذلك فان دالة التوزيع الاحتمالي المشترك ستكون كما يلي :

وهي عائلة اسية بـ k من المعلمات . فان:

$$f(x_1,....,x_n|\eta)=[\prod_{i=1}^{n}h(x)][c^*(\eta)]^n\exp[\eta_1\sum_{i=1}^{n}t_1(x_i)+......+\eta_k\sum_{i=1}^{k}t_k(x_i)]$$

$$T(X)=\sum_{i=1}^{n}t_1(x_i),......,\eta_k\sum_{i=1}^{k}t_k(x_i)$$

هي احصاءة كافية لـ η وذلك باستخدام النظرية العاملية فنحصل على :

$$f(x_1,....,x_n|\eta)=\underbrace{[\prod_{i=1}^{n}h(x)]}_{h(x)}\underbrace{[c^*(\eta)]^n\exp[\eta_1\sum_{i=1}^{n}t_1(x_i)+......+\eta_k\sum_{i=1}^{k}t_k(x_i)]}_{g(T(X)|\eta)}$$

ب- <u>نظرية ليمان – شيفي لاصغر احصاءة كافية</u> :

Lehman-Scheffé Theorem for Minimal Sufficient Statistics

افرض ان $f(x|\theta)$ هي دالة كتلة او كثافة احتمالية للعينة X وان $C_{x,y}$ هـو ثابـت . وافـرض ان هنـاك دالـة اخـرى T(x). ان النسـبة $R_{x,y}(\theta)=\frac{f(x|\theta)}{f(y|\theta)}=C_{x,y}$ لجميع قيم θ في Θ . اذا كان T(x)=T(y) فان T(x) هي اصغر احصاءة كافية لـ θ ، انظر (Olive,2006) .

ولبرهنة ان T(x) هي اصغر احصاءة كافية لـ η اذا كانت $\eta_1,.......,\eta_n$ لاتحقق شرط الخطية، انظر (Johanson (1979,p.3) و (Lehman (1983,p.44) .

تقدير الامكان الاعظم Maximum Likelihood Estimation

افرض ان $f(x|\theta)$ هي دالة كتلة (كثافة) احتمالية للعينة X ، فان دالة الامكان هي $L(\theta)=f(x|\theta)$ لكل $X=(X_1,X_2.......,X_n)$. ان قيمة المعلمة $\hat{\theta}(X)\in\Theta$ هي التي تجعل $L(\theta|X)$ معظمة عند ثبات X. لذلك فان (MLE) للمعلمة θ والتي تعتمد على العينة X هي $\hat{\theta}(X)$، للمزيد انظر (Olive,2005) .

ان من الخواص المهمة في تقدير الامكان الاعظم والتي تكون مهمة بالنسبة للعائلة الاسية هي خاصية اللاتغير (Invariance Property) والتي تنص على :

اذا كان $\hat{\theta}$ هو تقدير امكان اعظم للمعلمة θ بفضاء المعلمات Ω ، فان $h(\hat{\theta})$ هي تقدير امكان اعظم لـ $h(\theta)$ ، اي ان (MLE) ثابت عندما تكون الدالة متباينة .

وللتعرف على خاصية اللاتغير نفرض ان $\theta^*=h(\theta)$ وعليه فان $\theta=h^{-1}(\theta^*)$ لذلك فان :

$$L(\theta|X_1,X_2,...,X_n)=L(h^{-1}(\theta^*)|X_1,X_2,...,X_n)$$

والتي يمكن الرمز لها بـ $L^*(\theta^*|X_1,X_2,...,X_n)$ ، وبالتالي فان تقدير الامكان الاعظم لـ θ (اذا كان موجودا) يكون كما يلي

$$MaxL(\theta|X_1,X_2,...,X_n)=MaxL^*(\theta^*|X_1,X_2,...,X_n)$$

حيث ان $\theta^*\in\Omega^*$ وان Ω^* هي فضاء المعلمة المتحولة، للمزيد انظر (عبد المجيد حمزة الناصر، ١٩٨٨) .

المقدر غير المتحيز ذي التباين القليل (UMVUE) والحد الادنى لمتباينة فريشيه – راو كرامر (FCRLB)

افرض ان $R(\theta)$ هي دالة حقيقية لـ θ وان $W=W(X_1,X_2.......,X_n)$ هو مقدر لـ $R(\theta)$ فان الجزء المتحيز للمقدر W لـ $R(\theta)$ هو

$$b(R(\theta))=E(W)-R(\theta)$$

وان متوسط مربعات الاخطاء (mean square error) للمقدر W لـ $R(\theta)$ هو

$$Mse(R(\theta))=E[W-R(\theta)]^2=Var(W)+[b(R(\theta))]^2 \qquad (6)$$

ان W هو مقدر غير متحيز لـ $R(\theta)$ اذا كان $W=R(\theta)$ لجميع قيم $\theta\in\Theta$ وان W_U هو (UMVUE) لـ $R(\theta)$ (اذا كان W_U مقدر غير متحيز لـ $R(\theta)$ وان $Var(W_U)\le Var(W)$ لجميع قيم θ حيث ان W اي مقدر غير متحيز لـ $R(\theta)$) .

وللتعـرف علـى ماهيـة (FCRLB) او متباينـة المعلومـات ، نفـرض ان $X_1,X_2.......,X_n$ هي متغيرات عشوائية مستقلة بدالة كتلة (كثافة) احتماليـة مشـتركة $f(x|\theta)$ والتـي تحقـق المعادلـة (6) . افـرض ان $W(X_1,X_2.......,X_n)$ هو اي مقدر لـ $R(\theta)=E(W)$ فان

$$VarW(X)\ge\frac{[\frac{dE[W(X)]}{d\theta}]^2}{E[\frac{\partial}{\partial\theta}\log f(X|\theta)]^2}=\frac{[R'(\theta)]^2}{I_n(\theta)}$$

ان المقدار $FCRLB=\frac{[R'(\theta)]^2}{I_n(\theta)}$ هو الحد الادنى لمتباينة فريشيه – راو كرامر للمقدرات غير المتحيزة ذي التباين القليل لـ $R(\theta)$ ، للمزيد انظر (Wijsman(1973) و Joshi(1976)) .

نظرية الغاية المركزية لمعلمة واحدة من التوزيعات التي تتبع العائلة الاسية :

ان العالمين Cox و Hinkley لاحظا في حالة العائلة الاسية المنتظمة بمعلمة واحدة ان $T_n=\frac{1}{n}\sum_{i=1}^{n}t(X_i)$ هو مقدر غير متحيز منتظم ذو تباين قليل (UMVUE) وان مقدر الامكان الاعظم (MLE) لتوقع هذا المقدار هو $\mu_T=E_\theta(T_n)=E_\theta[t(X)]$ ، $\sigma_T^2=Var_\theta[t(X)]$. ان هذه القيم يمكن ايجادها باستخدام صيغ Casella و Berger وكما يلي:

$$\mu_T=\frac{-c'(\theta)}{c(\theta)w'(\theta)}=\frac{-\partial}{\partial\eta}\ln(c^*(\eta)), \qquad (7)$$

$$\sigma_T^2=\frac{\frac{-\partial^2}{\partial\theta^2}\ln(c(\theta))-[w''(\theta)]\mu_T}{[w'(\theta)]^2}=\frac{-\partial^2}{\partial\eta^2}\ln(c^*(\eta)) \qquad (8)$$

اذا كان $\theta=g(\eta)$ وان $g'(\eta)\neq0$ فان $I_1(\theta)=\frac{I_1(\eta)}{[g'(\eta)]^2}$. هذا هو التعريف المناسب عند استخدام طريقة دلتا او عندما تكون الدالة g متباينة وشاملة .

نظريـــة (٢) : افـرض ان $x_1,.......,x_n$ هـو توزيعـات عشـوائية مسـتقلة و متشـابهة مـن عائلـة اسـية بمعلمـة واحـدة بــ$E(t(X)) = \mu_T \equiv g(\eta)$ و $Var(T(X)) = \sigma_T^2$. فان

١- اذا كـان

$$\sqrt{n}[T_n - \mu_T] \sim N(0, I_1(\eta))$$

حيث ان

$$I_1(\eta) = \sigma_T^2 = g'(\eta) = \frac{[g'(\eta)]^2}{I_1(\eta)}$$

فان T_n هي كفوءة محاذيا .

٢- اذا كان $\eta = g^{-1}(\mu_T), \hat{\eta} = g^{-1}(T_n)$ وان $g^{-1'}(\mu_T) \neq 0$ فان

$$\sqrt{n}(\hat{\eta} - \eta) \sim N(0, \frac{1}{I_1(\eta)})$$

٣- اذا كـان $\theta = w^{-1}(\eta), \hat{\theta} = w^{-1}(\hat{\eta})$ وان $w^{-1'}(\eta) \neq 0$

فان

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \frac{1}{I_1(\theta)})$$

البرهان :

١- من نظرية الغاية المركزية $I_1(\eta) = \sigma_T^2 = g'(\eta)$ فان

$$\ln(f(x|\eta)) = \ln(h(x)) + \ln(c^*(\eta)) + \eta t(x)$$

$$\frac{\partial}{\partial \eta}\ln(f(x|\eta)) = \frac{\partial}{\partial \eta}\ln(c^*(\eta)) + t(x) = -g(\eta) + t(x)$$

$$\frac{\partial^2}{\partial \eta^2}\ln(f(x|\eta)) = \frac{\partial^2}{\partial \eta^2}\ln(c^*(\eta)) = -g'(\eta)$$

$$\therefore I_1(\eta) = -E[\frac{\partial^2}{\partial \eta^2}\ln(f(x|\eta))] = \sigma_T^2 = g'(\eta)$$

٢- باستخدام طريقة دلتا

$$\because g(\eta) = \mu_T \Rightarrow \eta = g^{-1}(\mu_T) \text{ and } \hat{\eta} = g^{-1}(\hat{\mu}_T)$$

فان

$$\sqrt{n}(g^{-1}(\hat{\mu}_T) - g^{-1}(\mu_T)) \sim N(0, \sigma_T^2[g^{-1'}(\mu_T)]^2)$$

وان

$$g^{-1'}(\mu_T) = \frac{1}{g'(g^{-1}(\mu_T))} = \frac{1}{g'(\eta)}$$

كذلك

$$\sigma_T^2[g^{-1'}(\mu_T)]^2 = \frac{[g'(\eta)]^2}{I_1(\eta)} \cdot \frac{1}{[g'(\eta)]^2} = \frac{1}{I_1(\eta)}$$

$$\Rightarrow \sqrt{n}(\hat{\eta} - \eta) \sim N(0, \frac{1}{I_1(\eta)})$$

٣- باستخدام طريقة دلتا

$$\because \theta = w^{-1}(\eta), \hat{\theta} = w^{-1}(\hat{\eta})$$

فان

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \frac{[w^{-1'}(\eta)]^2}{I_1(\eta)})$$

وان

$$\frac{[w^{-1'}(\eta)]^2}{I_1(\eta)} = \frac{1}{I_1(\theta)} .$$

$$\therefore \sqrt{n}(\hat{\theta} - \theta) \sim N(0, \frac{1}{I_1(\theta)})$$

لــــذا فــــان T_n هـــو (UMVUE) لــــ μ_T ، وللمزيــــد انظـــر (Olive,D.J,2006).

نظريـة الغايـة المركزيـة لـ K من المعلمـات للتوزيعات التـي تتبـع العائلة الاسية :

افرض ان $X_1,.......,X_n$ هي متغيرات عشوائية مستقلة ولها نفس التوزيع مـــن K مـــن المعلمـــات ، فـــان اصـــغر احصـــاءة كافيـــة هـــي $T_n = \frac{1}{n}(\sum_{i=1}^n t_1(X_i),............,\sum_{i=1}^n t_k(X_i))$.

افرض ان $\mu_T = (E(t_1(X)),........E(t_k(X)))$.

انظر (Lehmann (1986, p. 66 و Lehmann (1999, pp. 497, 499)

عندما $\eta \in \Omega$ فان

$$E(t_i(X)) = \frac{-\partial}{\partial \eta_i}\ln(c^*(\eta)) \; ;$$

$$Cov(t_i(X), t_j(X)) = \sigma_{ij} = \frac{-\partial^2}{\partial \eta_i \partial \eta_j}\ln(c^*(\eta))$$ وان مصـــــــفوفة المعلومات هي $I(\eta) = [I_{ij}]$

حيث ان

$$I_{ij} = E[\frac{\partial}{\partial \eta_i}\ln(f(x|\eta))\frac{\partial}{\partial \eta_j}\ln(f(x|\eta))] = -E[\frac{\partial^2}{\partial \eta_i \partial \eta_j}\ln(f(x|\eta))]$$

نظريـة (٣) : اذا كـان $X_1,.......,X_n$ هـي متغيرات عشوائية مستقلة ولهـا نفس التوزيع من العائلة الاسية المنتظمة بـ K من المعلمات ، فان

$$\sqrt{n}(T_n - \mu_T) \sim N_k(0, I(\eta))$$

البرهان :

باستخدام نظرية الغاية المركزية متعددة المتغيرات

$$\sqrt{n}(T_n - \mu_T) \sim N_k(0, \Sigma)$$

حيث ان $\Sigma = [\sigma_{ij}]$ ، $\sigma_{ij} = I_{ij}$. فان

$$\ln(f(x|\eta)) = \ln(h(x)) + \ln(c^*(\eta)) + \sum_{i=1}^k \eta_i t_i(x)$$

$$\frac{\partial}{\partial \eta_i}\ln(f(x|\eta)) = \frac{\partial}{\partial \eta_i}\ln(c^*(\eta)) + t_i(x)$$

$$I_{ij} = -E[\frac{\partial^2}{\partial \eta_i \partial \eta_j}\ln(f(x|\eta))] = \frac{\partial^2}{\partial \eta_i \partial \eta_j}\ln(c^*(\eta)) = \sigma_{ij}$$

وللتعـرف اكثـر علـى نظريـة الغايـة المركزيـة متعـددة المتغيـرات انظـر (Olive,D.J,2006) .

المحاكاة :

لقد تم كتابـة برنـامج باسـتخدام البرنـامج الاحصـائي R/Splus لتوضـيح نظريـة الغايـة المركزيـة للبيانـات $X_1,.......,X_n$ والتـي لهـا توزيـع عشـوائي متشـابه مـن التوزيـع الطبيعـي ، توزيـع كامـا ، توزيـع ثنـائي الحـدين و توزيـع بواسـون . الدالـة تقـوم بتوليـد مجموعـة مـن البيانـات بحجـم n ويـتم حسـاب

ISSN : 1813 – 1662

$\bar{X}_1$ من هذه البيانات . هذه الخطوة تكرر ١٠٠ مرة . فيظهر لدينا المتجه $(\bar{X}_1, \bar{X}_2, ..., \bar{X}_n)$. المخططات لهذه الاوساط يجب ان تكون مشابهة لدالة التوزيع الطبيعي المتماثلة عندما تكون n كبيرة بشكل كافي. البرنامج كما يلي :

```
cltsim <- function(n=100, nruns=100){
xbar <- 1:nruns
for(i in 1:nruns){
xbar[i] <- mean(rgamma(n,1,3))}
list(xbar=xbar)}
z1 <- cltsim(n=1)
z5 <- cltsim(n=5)
z25 <- cltsim(n=25)
z50 <- cltsim(n=50)
z100 <- cltsim(n=100)
par(mfrow=c(2,3))
hist(z1$xbar)
hist(z5$xbar)
hist(z25$xbar)
hist(z50$xbar)
hist(z100$xbar)
```

حيث نقوم بتغيير الخطوة الرابعة في البرنامج فنحصل على الرسومات الموجودة في الصفحات (١٩–٢٢) من هذا البحث .

ان البيانات التي اجريت لها المحاكاة هي كما يلي :

جدول (١) : الأعداد العشوائية المسحوبة من التوزيع الطبيعي

4.471882	3.705662	7.582331	9.675885
7.152609	7.203354	4.865256	2.705655
4.219924	5.987487	3.403227	6.576371
4.881086	4.817792	5.803686	7.002315
5.791137	4.427378	5.089493	8.412478
5.990214	4.011861	8.921393	8.630030
7.395105	7.364920	6.365172	10.035282
7.091677	5.906635	6.564209	4.711267
8.287166	4.320424	4.970319	2.495526
7.639815	2.849497	5.539044	8.361741
4.843444	9.463603	7.186727	9.039826
7.944343	5.249753	3.943314	3.167953
10.634625	8.403656	6.247930	7.256564
7.512267	6.174204	3.853303	7.079254
3.176224	8.183669	7.574186	4.257860
6.072690	7.538367	3.642546	7.273978
5.277725	6.622952	6.469314	5.580786
6.962571	2.292401	4.979590	7.990239
8.247842	6.212651	7.497432	3.624177
4.937102	7.674392	8.894339	3.418695
6.476011	6.767703	8.824314	3.499203
6.145230	7.668136	6.316944	4.131584
5.435832	6.099950	7.646350	4.633205
3.649043	5.309545	4.353089	5.733341
5.391281	9.426679	8.096673	2.637288

جدول (٢) : الاعداد العشوائية المسحوبة من توزيع كاما

0.792493187	0.263108794	0.290888799	0.044650348
0.120987701	0.075149900	0.086788011	0.598838366
0.229075753	0.251844302	0.086467557	0.106473160
0.374806471	0.553578053	0.256317768	0.054877577
0.506029919	0.762606687	0.082367438	0.093387638
0.058410616	0.052031138	0.372874707	0.237716091
0.038156948	0.688199621	0.074384842	0.579116217
0.304255992	1.327930901	1.528399738	0.103693919
0.204190880	0.456989452	0.335651375	0.032588828
0.122858672	0.032247917	0.224669904	0.107487511
0.092327629	0.153587957	0.481194960	0.131904236
0.588492664	0.001599823	0.646408002	0.482208929
0.806400397	0.276445922	0.621755841	0.006995997
0.077792808	0.022603013	1.119509485	0.754112191
0.015498341	0.400792800	0.075386828	0.633363121
0.321607888	0.047925482	0.290491730	0.498935134
0.218185537	0.105957748	0.029907883	0.159792400
0.208438538	0.406111482	0.230146012	0.141583034
0.192301687	0.020830336	0.680771865	0.110880609
0.140180929	0.145958005	0.597880118	0.272237764
0.059208227	0.406750402	0.137489879	0.247603674
0.056636466	0.293105436	0.931679219	0.476495053
0.271626734	0.043735458	0.872267369	0.072444185
0.821428820	0.438683801	0.015441771	0.117944394
0.606849452	0.503883471	0.428739308	0.352791323

جدول (٣) : الاعداد العشوائية المسحوبة من توزيع ثنائي الحدين

4	3	4	3	3	2	6	3	3	6
3	4	4	4	2	5	3	6	5	4
3	3	5	3	3	2	5	5	3	4
2	5	1	3	2	3	3	3	1	2
0	5	3	3	3	6	4	5	3	2
4	1	3	4	3	3	4	4	4	4
3	3	3	4	3	4	2	3	3	3
2	4	3	0	4	4	3	4	5	3
4	1	5	0	5	4	1	4	3	3
4	2	3	3	5	2	3	4	2	3

جدول (٤) : الاعداد العشوائية المسحوبة من توزيع بواسون

3	0	2	0	3	1	1	1	2	2
3	5	3	5	1	4	1	2	2	1
2	5	1	3	2	3	2	1	2	1
5	1	1	0	3	1	4	3	2	5
1	1	1	1	3	0	3	2	0	2
2	5	3	3	3	0	1	5	7	3
1	3	2	0	0	2	0	4	2	4
2	3	3	1	2	4	2	0	1	2
2	2	2	2	4	3	0	0	1	1
1	4	2	1	2	0	1	1	1	6

وكانت نتيجة المحاكاة للبيانات العشوائية المولدة اعلاه كما يلي :

جدول (٥) : الوسط الحسابي والانحراف المعياري للتوزيع الطبيعي بحجم عينة
(١ ، ٥ ، ٢٥ ، ٥٠ ، ١٠٠) على التوالي

Distribution	Sample Size	$\overline{X}$		Normal	
		mean	S.D.	mean	S.D.
Normal	1	5.879	2.035	6.000	2.000
	5	6.009	0.807	6.000	0.894
	25	5.987	0.396	6.000	0.400
	50	5.988	0.254	6.000	0.283
	100	6.004	0.193	6.000	0.200

نلاحظ من الجدول (٥) بانه كلما كبر حجم العينة $n \geq 30$ كلما اقترب الوسط الحسابي والانحراف المعياري (التقريبي) للاعداد العشوائية المسحوبة من التوزيع الطبيعي لـ $\overline{X}$ (وبحجم عينات مختلفة) من شكل التوزيع الطبيعي الحقيقي (Normal) .

جدول (٦) : الوسط الحسابي والانحراف المعياري لتوزيع كاما بحجم عينة
(١ ، ٥ ، ٢٥ ، ٥٠ ، ١٠٠) على التوالي

Distribution	Sample Size	$\overline{X}$		Normal	
		mean	S.D.	mean	S.D.
Gamma	1	2.882	3.221	3.000	3.000
	5	2.977	1.206	3.000	1.342
	25	2.979	0.591	3.000	0.600
	50	2.978	0.379	3.000	0.424
	100	3.005	0.292	3.000	0.300

نلاحظ من الجدول (٦) بانه كلما كبر حجم العينة $n \geq 30$ كلما اقترب الوسط الحسابي والانحراف المعياري (التقريبي) للاعداد العشوائية المسحوبة من توزيع كاما لـ $\overline{X}$ (وبحجم عينات مختلفة) من شكل التوزيع الطبيعي الحقيقي (Normal) .

جدول (٧) : الوسط الحسابي والانحراف المعياري لتوزيع ثنائي الحدين بحجم عينة
(١ ، ٥ ، ٢٥ ، ٥٠ ، ١٠٠) على التوالي

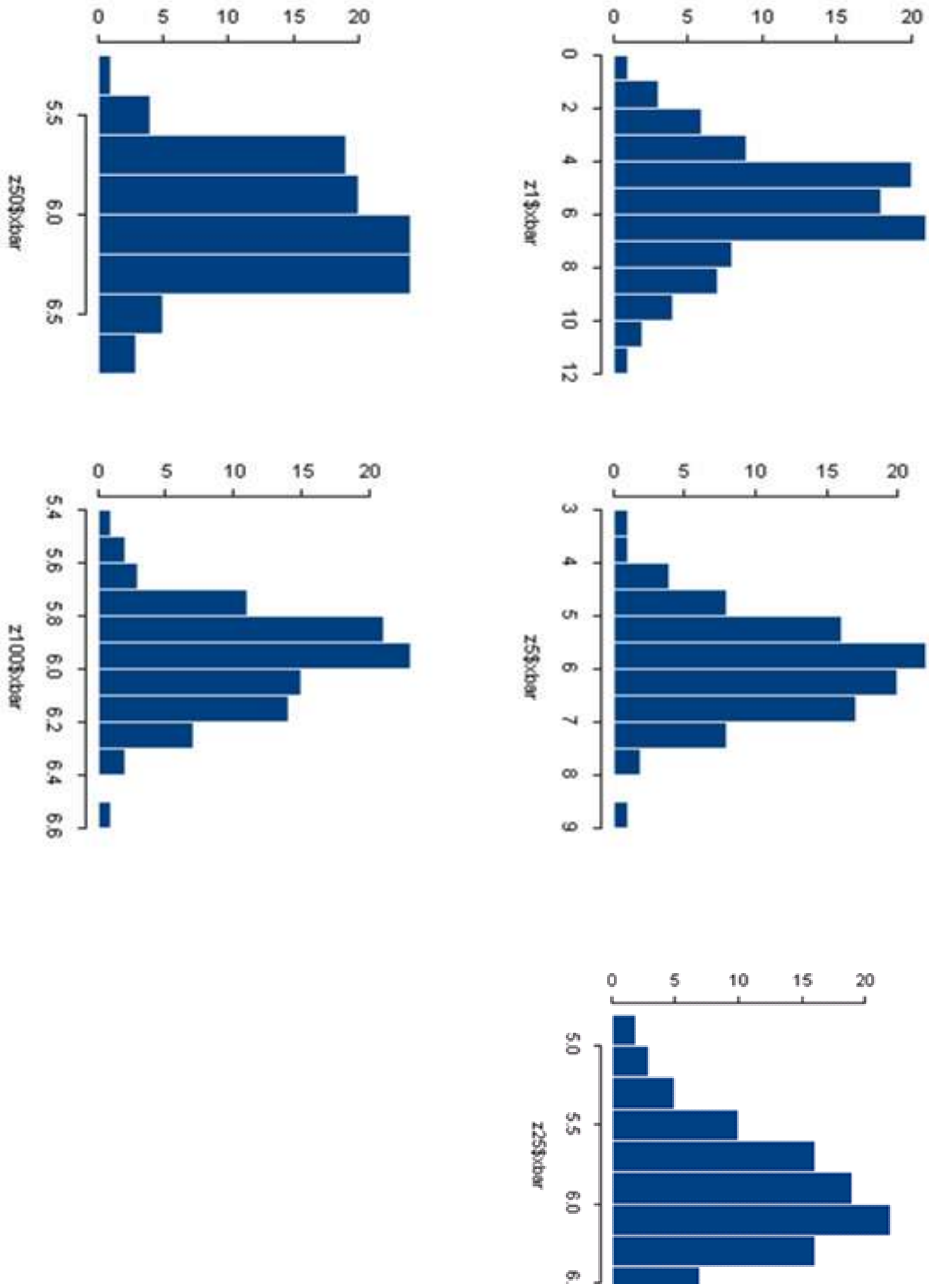
Distribution	Sample Size	$\overline{X}$		Normal	
		mean	S.D.	mean	S.D.
Binomial	1	3.255	1.322	3.200	1.386
	5	3.083	0.691	3.200	0.620
	25	3.201	0.261	3.200	0.277
	50	3.203	0.177	3.200	0.196
	100	3.205	0.132	3.200	0.139

نلاحظ من الجدول (٧) بانه كلما كبر حجم العينة $n \geq 30$ كلما اقترب الوسط الحسابي والانحراف المعياري (التقريبي) للاعداد العشوائية المسحوبة من توزيع ثنائي الحدين لـ $\overline{X}$ (وبحجم عينات مختلفة) من شكل التوزيع الطبيعي الحقيقي (Normal) .

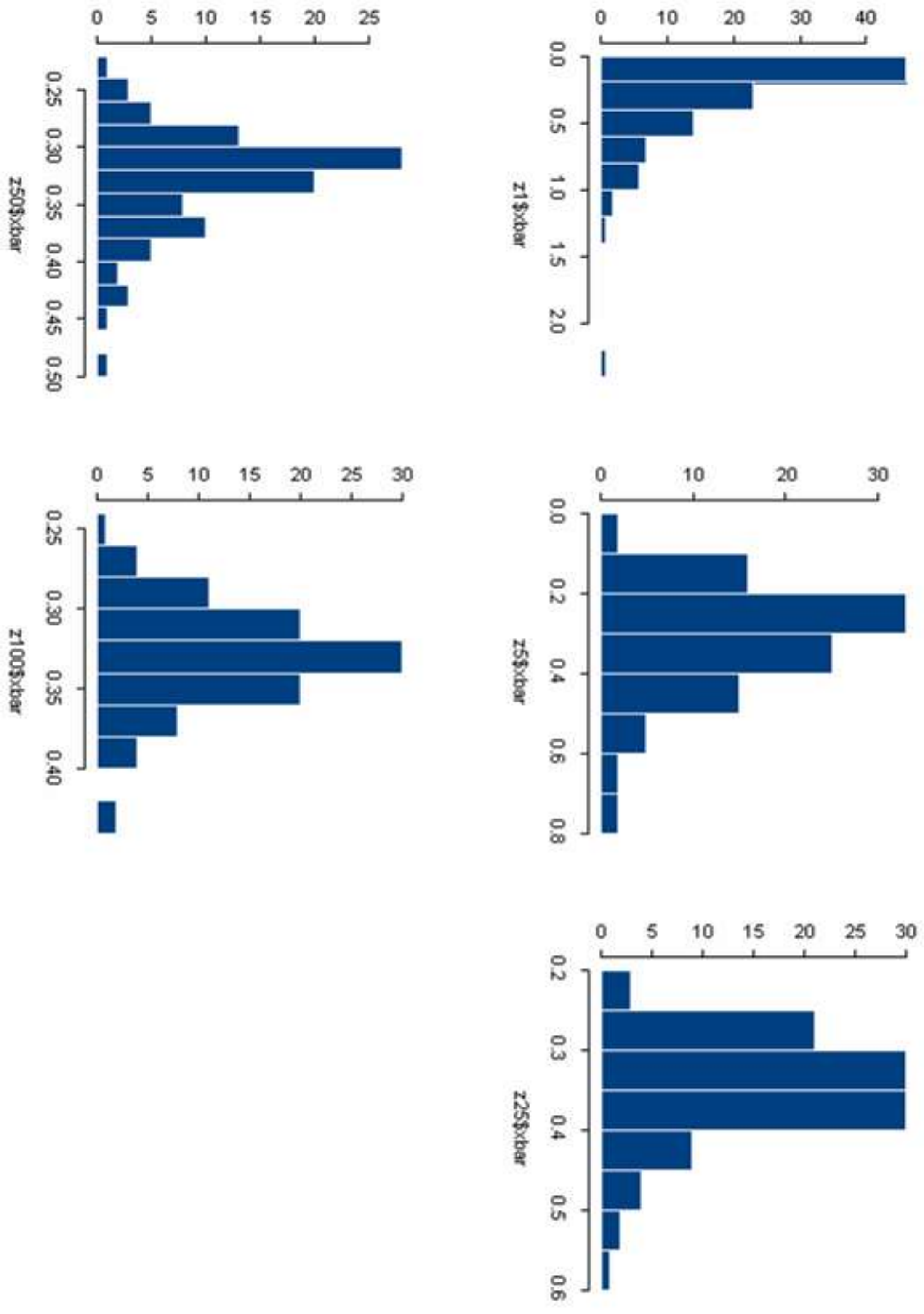
جدول (٨) : الوسط الحسابي والانحراف المعياري لتوزيع بواسون بحجم عينة
(١ ، ٥ ، ٢٥ ، ٥٠ ، ١٠٠) على التوالي

Distribution	Sample Size	$\overline{X}$		Normal	
		mean	S.D.	mean	S.D.
Poisson	1	2.805	1.806	3.000	1.732
	5	3.011	0.827	3.000	0.775
	25	2.973	0.325	3.000	0.346
	50	2.994	0.250	3.000	0.245
	100	3.007	0.150	3.000	0.173

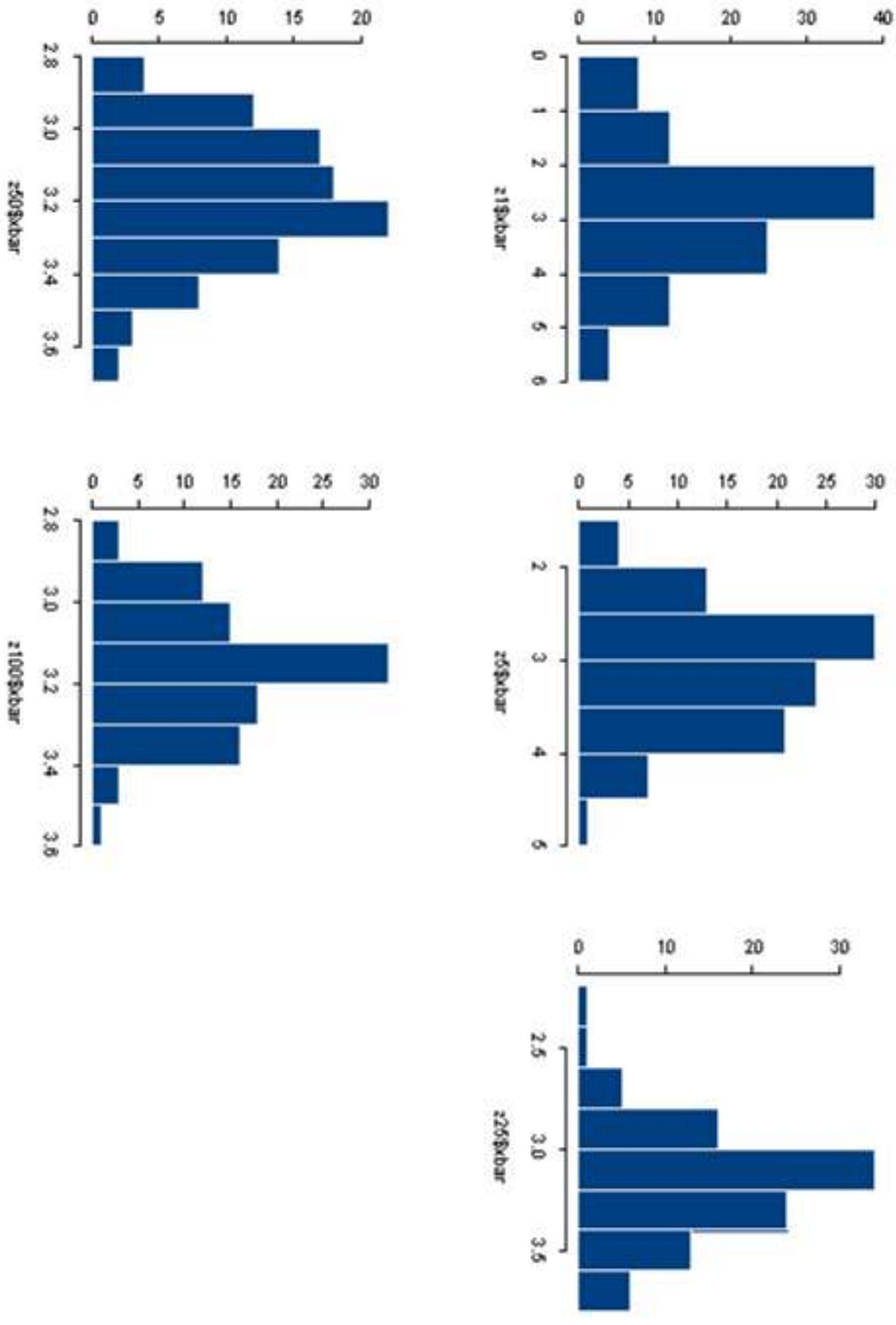
نلاحظ من الجدول (٨) بانه كلما كبرحجم العينة $n \geq 30$ كلما اقترب الوسط الحسابي والانحراف المعياري (التقريبي) للاعداد العشوائية المسحوبة من توزيع بواسون لـ $\overline{X}$ (وبحجم عينات مختلفة) من شكل التوزيع الطبيعي الحقيقي (Normal) .



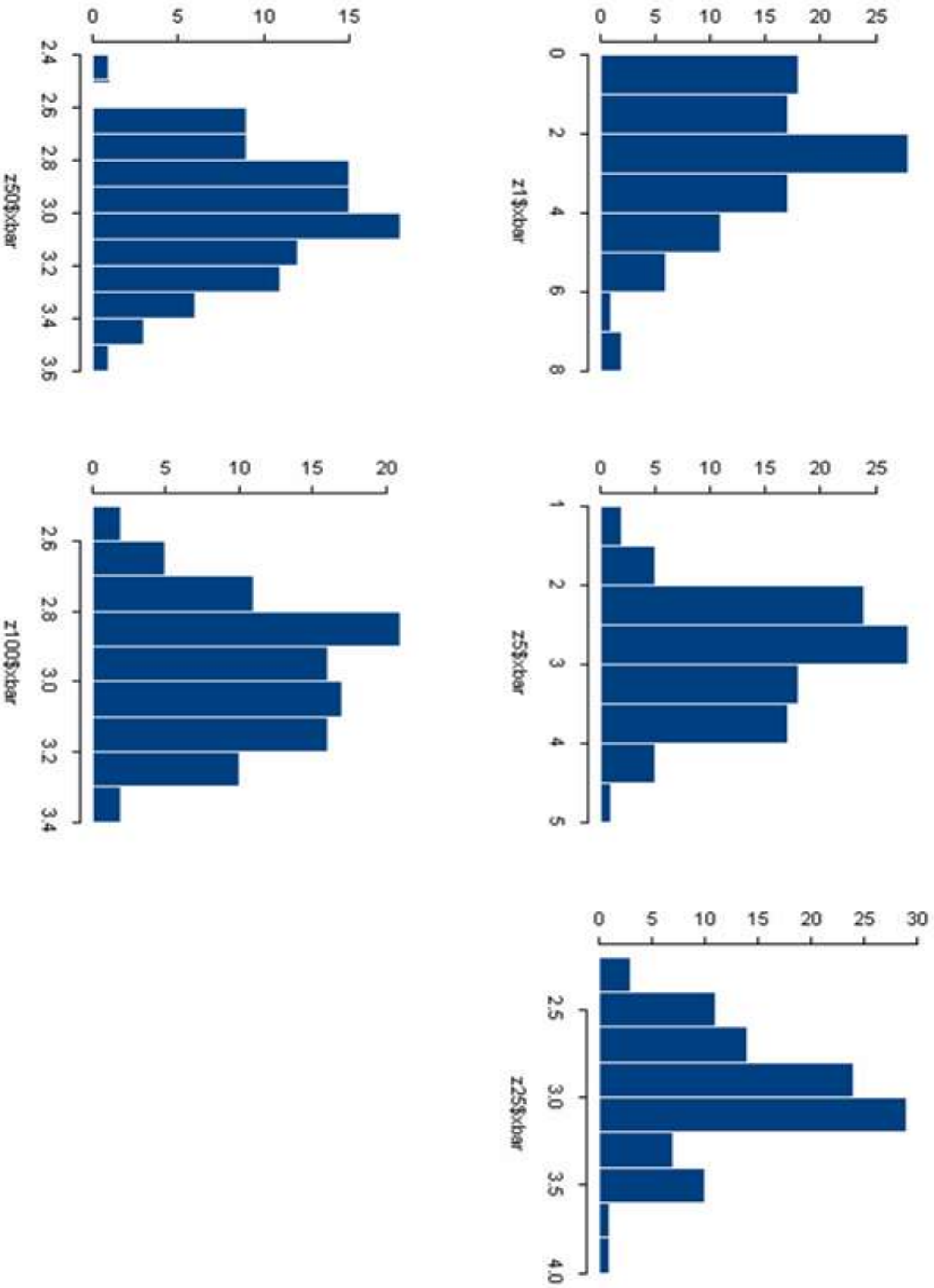
المدرج التكراري لتوزيع $\overline{X}$ في الجدول (٥)



المدرج التكراري لتوزيع $\overline{X}$ في الجدول (٦)



المدرج التكراري لتوزيع $\overline{X}$ في الجدول (7)



المدرج التكراري لتوزيع $\overline{X}$ في الجدول (٨)

١– نلاحظ من المدرج التكراري في الجدول (٥) المستخرج من البرنامج المذكور اعلاه ان توزيع المعاينة لـ $\overline{X}$ للتوزيع الطبيعي يمكن ان تؤول الى التوزيع الاحتمالي الطبيعي عندما يكون حجم العينة كبير $(n \geq 30)$.

٢– نلاحظ من المدرج التكراري في الجدول (٦) المستخرج من البرنامج المذكور اعلاه ان توزيع المعاينة لـ $\overline{X}$ لتوزيع كاما يمكن ان تؤول الى التوزيع الاحتمالي الطبيعي عندما يكون حجم العينة كبير $(n \geq 30)$.

٣– نلاحظ من المدرج التكراري في الجدول (٧) المستخرج من البرنامج المذكور اعلاه ان توزيع المعاينة لـ $\overline{X}$ لتوزيع ثنائي الحدين يمكن ان

تـؤول الـى التوزيـع الاحتمـالي الطبيعـي عنـدما يكـون حجـم العينـة كبيـر $(n \geq 30)$.

٤– نلاحظ من المدرج التكراري في الجدول (٨) المستخرج من البرنامج المذكور اعلاه ان توزيع المعاينة لـ $\overline{X}$ لتوزيع بواسون يمكن ان تؤول الى التوزيع الاحتمالي الطبيعي عندما يكون حجم العينة كبير $(n \geq 30)$.

الاستنتاجات والتوصيات :

١– ان نظرية الغاية المركزية تمكننا من معرفة توزيع المعاينة لـ $\overline{X}$ حتى في حالة عدم معرفة توزيع المتغير الاصلي .

٢– ان العينات التي يكون حجمها اكبر من ٣٠ مشاهدة $(n \geq 30)$ ، فان توزيع متوسط العينات يقترب بشكل معقول من التوزيع الطبيعي . التقريب يكون افضل كلما كبر حجم العينة .

٣– اذا كان المجتمع الاصلي يتبع التوزيع الطبيعي ، فان متوسط العينة سوف يتوزع طبيعيا لاي حجم عينة n .

٤– نوصي باستخدام العوائل الاسية لايجاد التوزيعات الاحتمالية الاولية المتالفة عند استخدام احصاء بيز .

لمصادر

1 – الناصر ، عبد المجيد حمزة و رشيد ، ظافر حسين (١٩٨٨) . " الاستدلال الاحصائي ".

2- Barndorff-Nielsen, O. (1982)," Exponential Families " , in Encyclopedia of Statistical Sciences, eds. Kotz, S. and Johnson, N.L., John Wiley and Sons, NY, 587-596.

3- Casella, G., and Berger, R.L. (2002), " Statistical Inference " , Duxbury Press, 2nd ed., Belmont, CA.

4- Cox, D.R., and Hinkley, D.V. (1974), " Theoretical Statistics " , Chapman & Hall, London.

5- Johanson, S. (1979), " Introduction to the Theory of Regular Exponential Families", Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark.

6- Joshi, V.M. (1976),"On the Attainment of the Cramér-Rao Lower Bound, The Annals of Statistics, 4, 998-1002.

7- Le Cam, L. (1986), "The Central Limit Theorem around 1935 " , Statistical Science 1(1): 78-96.

8- Lehmann, E.L. (1983), " Theory of Point Estimation " , 1st ed., John Wiley and Sons, NY.

9- Lehmann, E.L. (1986), " Testing Statistical Hypotheses " , 2nd ed., John Wiley and Sons, NY.

10- Lehmann, E.L. (1999), " Elements of Large–Sample Theory " , Springer-Verlag, NY.

11- Olive, D.J. (2005), " Using Exponential Families in an Inference Course " Department of Mathematics , Southern Illinois University, USA.

12- Olive, D.J. (2006), " A Course in Statistical Theory " Department of Mathematics, Southern Illinois University, USA.

13- Schervish, M.J. (1995), " Theory of Statistics", Springer -Verlag, NY.

14- Wackerly, D., Mendenhall, W., and Scheaffer, R.(2002), " Mathematical Statistics with Applications " , 6th ed., Duxbury, 2002.

15- Wijsman, R.A. (1973), " On the Attainment of the Cramér-Rao Lower Bound, The Annals of Statistics, 1, 538-542.

Simulation Study for Central Limit Theorem for Exponential Families

(Received 27 / 10 / 2008 , Accepted 12 / 4 / 2009)

Abstract

This paper shows that under regularity conditions and if $Y_1,.......,Y_n$ are iid from a k–parameter regular exponential family with complete sufficient statistic T_n with $E(T_n) = \mu_T$, then $\sqrt{n}(T_n - \mu_T) \sim N_k(0, I(\eta))$ where $I(\eta)$ is the information matrix of the natural parameterization of the family. And also we can use the delta method in finding the Expectation and Variance for the complex function . In this paper, I generate random numbers for 2 continuous distributions and 2 discrete distributions which belong to the Exponential Family, the result of Simulation appears that the Central Limit Theorem has a good approximation for the Normal distribution .