جامعة دمشق - كلية العلوم

قسم الرياضيات

مقرر الإحصاء الرياضي

السنة الثالثة

الفصل الثاني للعام الدراسي 2013 – 2014

د. سلطان محمد الصلخدي

مراجعة عامة

بعض أهم التوزيعات الاحتمالية

1 – توزيع برنولي :

نقول عن متغير عشوائي منفصل X أنّه يتوزع وفق توزيع برنولي اذا كانت كثافته الاحتمالية معرفة بالشكل :

x	0	1
$P(X=x)$	q	p

مع العلم أنّ $0<p<1 \quad ; \quad p+q=1$ أي حالة تجربة ثنائية مكررة $n=1$ مرة, ولهذا التوزيع شكل أخر هو: $0<p<1 \quad \& \quad x=0,1 \quad , \mathrm{f}(x)=p^x(1-p)^{1-x}$

مبرهنة: إذا كان للمتغير X توزيع برنولي وسيطه p عندئذ فإنّ :

$$\mathrm{E(X)}=p \quad ; \quad \mathrm{Var(X)}=p(1-p)$$

2 – التوزيع الحداني:

نقول عن متغير عشوائي منفصل X أنّه يتوزع وفق التوزيع الحداني وسيطاه n و p إذا كانت كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \quad , \quad x=0,1,2,\dots,n \;\; ; \; 0<p<1 \;\; ,p+q=1$$

ونكتب رمزا على أنّ X يتوزع وفق التوزيع الحداني $b(x;n,p)$.

أنّ التوزيع الحداني يعالج التجارب الحدانية أو التجارب الثنائية وتعرف التجارب الثنائية بأنّها التجارب المكونة من n محاولة من المحاولات المستقلة وكل محاولة لها احتمالان يسمى الأول نجاح p ويسمى الثاني فشل q بحيث يبقى احتمال النجاح ثابتاً من محاولة لأخرى و يدل X على عدد مرات النجاح.

ملاحظة 1: لنفرض أنّ $\mathrm{X_1,X_2,\dots\dots,X_n}$ عينة عشوائية لمتغير عشوائي برنولي X وسيطه p فإنّ توزيع المتغير $\mathrm{Y}=\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathrm{X_i}$ هو متغير حداني وسيطاه n و p .

ملاحظة 2: لتكن $\mathrm{X_1,X_2,\dots\dots,X_\kappa}$ عينة عشوائية لمتغير حداني وسيطاه n و p فإنّ توزيع المتغير $\mathrm{Y}=\sum_{\mathrm{i=1}}^{\kappa}\mathrm{X_i}$ هو متغير حداني وسيطاه $n\kappa$ و p .

مبرهنة: إذا كان X متغير عشوائي له التوزيع الحداني وسيطاه n و p عندئذ فإنّ :

$$\mathrm{E(X)}=np \quad , \quad \mathrm{Var(X)}=np(1-p)$$

3 – التوزيع البواسوني:

نقول عن متغير عشوائي منفصل X أنّه يتوزع وفق التوزيع البواسوني وسيطه θ اذا كانت كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x) = \frac{\theta^x}{x!}\mathrm{e}^{-\theta}\,, \quad x = 0,1,\dots \qquad \theta = np > 0 \quad ; \quad 0 < p < 1$$

وهذا التوزيع هو نهاية لدالة الاحتمال للتوزيع الحداني وذلك عندما تصبح n كبيرة بقدر كاف و يصبح p صغير جداً بحيث يبقى الجداء مساويا عدداً ثابتاً.

ملاحظة 3: لتكن $X_1, X_2, \dots\dots, X_n$ عيّنة عشوائية لمتغير عشوائي بواسّوني وسيطه θ فإنّ توزيع المتغير $Y = \sum_{i=1}^{n} X_i$ هو متغير بواسّوني وسيطه $n\theta$.

مبرهنة: إذا كان X متغير عشوائي بواسوني وسيطه ϑ عندئذ فإنّ :

$$E(X) = \theta \quad ; \quad Var(X) = \theta$$

4 – التوزيع الهندسي:

نقول عن متغير عشوائي منفصل X أنّه يتوزع وفق التوزيع الهندسي وسيطه p اذا كانت كثافته الاحتمالية معرفة بالشكل:

$$f(x) = p(X = x) = p(1-p)^x \;,\; x = 0,1,2,\dots\dots, \;\; p + q = 1 \;\; ; \;\; 0 < p < 1$$

أو بالشكل:

$$f(x) = p(X = x) = p(1-p)^{x-1} \;, x = 1,2,3,\dots\dots, \;\; p + q = 1 \;\; ; \;\; 0 < p < 1$$

مبرهنة: إذا كان X متغير عشوائي له التوزيع الهندسي وسيطه p عندئذ يكون:

$$E(X) = \frac{q}{p} \qquad Var(X) = \frac{q}{p^2}$$

وفي الحالة الثانية:

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{q}{p^2}$$

5 – توزيع باسكال:

نقول عن متغير عشوائي منفصل X أنّه يتوزع وفق توزيع باسكال إذا كانت كثافته الاحتمالية معطاة بالشكل التالي :

$$f(x) = p(X = x) = \binom{x-1}{k-1}p^k(1-p)^{x-k} \;,\; x = k, k+1, k+2, \dots$$

6 – التوزيع المنتظم على الفترة $[a, b]$:

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق التوزيع المنتظم على الفترة $[a, b]$ إذا كانت كثافته الاحتمالية معطاة بالشكل التالي :

$$f(x)=\frac{1}{\mathrm{b-a}} \quad ; \quad \mathrm{a}\le x\le b$$

مبرهنة : إذا كان X متغيراَ عشوائيّاَ له التوزيع المنتظم على الفترة $[\mathrm{a,b}]$ عندئذ يكون:

$$E(X)=\frac{b+a}{2} \quad ; \quad \mathrm{Var(X)}=\frac{(\mathrm{b-a})^2}{12}$$

7 – التوزيع الطبيعي:

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق التوزيع الطبيعي وسيطاه σ^2, μ اذا كانت كثافته الاحتمالية معرفة بالشكل التالي :

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ;x\in\mathrm{R} \;\; ;\sigma>0 \;\; ; \;\; \mu\in R$$

ويكتب هذا التوزيع بالشكل $\mathrm{N}(\mu,\sigma^2)$.

مبرهنة : إذا كان X متغيراَ عشوائيّاَ طبيعيّاَ فعندئذ يكون : $\mathrm{Var(X)}=\sigma^2 \quad ; \quad \mathrm{E(X)}=\mu$

ملاحظة 4 : إذا كان $\mu=0,\sigma=1$ فإنّ هذا التوزيع يؤول إلى التوزيع الطبيعي المعياري وتصبح دالة الكثافة الاحتمالية بالشكل : $f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \;;x\in\mathrm{R}$

ونرمز له في هذه الحالة بالرمز $\mathrm{N}(0,1)$ ويكون هنا التوقع والتباين معطيأنّ كما يلي :

$$\mathrm{E(X)}=0 \quad , \quad \mathrm{Var(X)}=1$$

ونظرًا لأهمية الّتوزيع الطبيعي المعياري نرمز للمتغير بالرمز Z وتصبح دالة الكثافة بالشكل التالي :

$$f(z)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\mathrm{z}^2}{2}} \quad ; \quad \mathrm{z}\in\mathrm{R}$$

ملاحظة 5 : إذا كانت المتغيرات العشوائية $\mathrm{X_1,X_2,\ldots\ldots,X_n}$ متغيرات مستقلة وكانت توزيعاتها هي توزيعات طبيعية $X_{\mathrm{i}}\sim\mathrm{N}(\mu_{\mathrm{i}},\sigma_{\mathrm{i}}^2)$ ($i=1,2,\ldots,n$) فإنّ توزيع المتغّير العشوائي $Y=X_1+X_2+\cdots.+X_n$ هو التوزيع الطبيعي وسيطاه $\sum_{i=1}^{n}\sigma_i^2$ ، $\sum_{i=1}^{n}\mu_i$.

ملاحظة 6 : إذا كانت هذه المتغيرات العشوائية $\mathrm{X_1,X_2,\ldots\ldots,X_n}$ تشكل عينة عشوائية للمتغير العشوائي X الذي له التوزيع الطبيعي $N(\mu,\sigma^2)$ فإنّ توزيع المتغّير العشوائي $Y=X_1+X_2+\cdots.+X_n$ هو التوزيع الطبيعي وسيطاه $n\sigma^2$, $n\mu$.

ملاحظة 7 : إذا كان $\bar{X}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathrm{X_i}$ هو متوسط العينة العشوائية فإنّ : $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$.

8 – التوزيع الأسي :

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق التوزيع الأسي وسيطه θ إذا كانت كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x) = \theta e^{-\theta x} \qquad ; \quad x > 0 \quad \& \quad \theta > 0$$

مبرهنة: إذا كانX يتوزع وفق التوزيع الأسي وسيطه θ عندئذ يكون:

$$E(X) = \frac{1}{\theta} \qquad , \qquad Var(X) = \frac{1}{\theta^2}$$

ملاحظة 8 : هناك شكل أخر لدالة الكثافة:

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق التوزيع الأسي وسيطه α إذا كانت كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}} \qquad ; \quad x > 0 \quad \& \quad \alpha > 0$$

وعندئذ يكون لدينا :

$$E(X) = \alpha \qquad ; \qquad Var(X) = \alpha^2$$

9 – توزيع غاما وسيطه α:

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق توزيع غاما وسيطه α إذا كانت كثافته الاحتمالية معرفة بالشكل التالي :

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \quad , x > 0 \;; \; \alpha > 0$$

حيث :

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx \;, \;\; \Gamma(\alpha) = (\alpha-1)(\alpha-2)\Gamma(\alpha-2)$$

ومن أجل كون n صحيحاً موجباً و $\alpha = n$ نجد:

$$\Gamma(n) = (n-1)! = (n-1)(n-2)(n-3)\dots 3.2.1$$

وفي حالة كون $\alpha > 0$ حقيقياً موجباً وليست صحيحة فإنّنا نهتم بالقيمة $\Gamma\left(\frac{1}{2}\right)$ حيث يمكننا أنّ نثبت أنّ:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

10 – توزيع غاما بوسيطين α, β :

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق توزيع غاما وسيطاه α, β إذا كانت كثافته الاحتمالية معرفة بالشكل :

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad , \quad x > 0 \quad , \quad \alpha, \beta > 0$$

مبرهنة: إذا كان X متغير عشوائي يخضع لتوزيع غاما وسيطاه α, β فعندئذ يكون لدينا :

$$E(X) = \alpha\beta \qquad Var(X) = \alpha\beta^2$$

ملاحظة 9: هناك شكل أخر لدالة الكثافة:

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق توزيع غاما وسيطاه $\alpha, \beta = \frac{1}{\lambda}$ إذا كانت كثافته الاحتمالية معرفة بالشكل :

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad , \quad x > 0 \quad , \quad \alpha, \beta = \frac{1}{\lambda} > 0$$

وفي هذه الحالة يكون لدينا : $Var(X) = \frac{\alpha}{\lambda^2}$, $E(X) = \frac{\alpha}{\lambda}$.

ملاحظة 10: إذا كان $\alpha = 1 \, , \, \beta > 0$ فإنّ توزيع غاما والذي وسيطاه $\alpha = 1$ و β يؤول إلى التوزيع الآسي وسيطه $\lambda = \frac{1}{\beta}$ أي أنّ كثافته تصبح بالشكل التالي $G(1,\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$.

ومن أجل κ صحيح موجب وبفرض أنّ $\beta = 2$, $\alpha = \frac{\kappa}{2}$ عندها يصبح $G(\alpha,\beta) = G\left(\frac{k}{2},2\right)$.

وأنّ الكثافة تصبح بالشكل : $\mathrm{G}\left(\frac{\kappa}{2},2\right) = \frac{1}{(2)^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} (x)^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}}$.

حيث أنّ $x > 0$ وهي كثافة توزيع كاي مربع درجة حريته k .

ملاحظة 11 : لتكن $X_1, X_2, \ldots \ldots, X_n$ عينة عشوائية لمتغير عشوائي غمّاوي والذي وسيطاه α, β فإنّ توزيع المتغير $\mathrm{Y} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}$ هو توزيع غمّاوي وسيطاه β, $n\alpha$.

11 – توزيع كاي مربّع $\chi^2_{(\mathrm{n})}$ درجة حريته n :

نقول عن متغير عشوائي مستمر X أنّ له توزيع كاي مربع درجة حريته n إذا كانت كثافته الاحتمالية معرفة بالشكل:

$$f(x) = \frac{1}{(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad , \; x \geq 0$$

مبرهنة : إذا كان X متغيراً عشوائياً يتوزع وفق توزيع كاي مربع درجة حريته n عندئذ يكون :

$$E(X) = n \qquad ; \qquad Var(X) = 2n$$

ملاحظة 11 : إذا كانت $X_1, X_2, \ldots \ldots, X_n$ كانت متغيرات عشوائية مستقلة من النمط كاي مربع درجة حريته الواحد فإنّ المتغير $Y = X_1 + X_2 + \cdots . + X_n$ له النمط كاي مربع درجة حريته n .

ملاحظة 12 : لتكن $\mathrm{X_1}, \mathrm{X_2}, \ldots \ldots, \mathrm{X_n}$ متغيرات عشوائية مستقلة تتوزع وفق توزيع كاي مربع بدرجات حرية هي على الترتيب $r_1, r_2, \ldots, r_n$ عندئذ يتوزع المتغير العشوائي $Y = X_1 + X_2 + \cdots . + X_n$ وفق توزيع كاي مربع درجة حريته $\sum_{i=1}^{n} r_i$.

ملاحظة 13 : ليكن $X \sim N(\mu, \sigma^2)$ عندئذ يتوزع $Z = \frac{X-\mu}{\sigma}$ وفق التوزيع الطبيعي المعياري ويتوزع أيضا Z^2 وفق توزيع كاي مربع درجة حريته الواحد .

12 – توزيع ستيودنت (t) درجة حريته n :

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق توزيع ستيودنت درجة حريته n اذا كانت كثافته الاحتمالية بالشكل التالي: $-\infty < x < +\infty$, $f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$

ملاحظة 14: إذا كان X متغير عشوائي طبيعي معياري أي أنّ $X\sim N(0,1)$ مستقل عن متغير عشوائي أخر له توزيع كاي مربع درجة حريته n أي أنّ $Y\sim\chi^2(n)$ فإنّ توزيع المتغير العشوائي $T=\frac{X}{\sqrt{\frac{Y}{n}}}$ هو توزيع ستيودنت درجة حريته n .

13 – توزيع فيشر درجتا حريته $\boldsymbol{m}$ و$\boldsymbol{n}$ على الترتيب :

نقول عن متغير عشوائي مستمر X أنّه يتوزع وفق فيشر درجتا حريته m وn على الترتيب اذا كانت كثافته الاحتمالية من الشكل :

$$f(x)=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}}(x)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}\quad ,\quad x>0$$

14 – توزيع بيتا :

نقول عن متغير عشوائي مستمر أنّه يتوزع وفق توزيع بيتا وسيطاه m,n إذا كانت دالة كثافته الاحتمالية معرفة بالشكل:

$$f(x)=\beta(m,n)=\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}x^{m-1}(1-x)^{n-1}\ ,\ m,n>0\ ,0<x<1$$

دالة بيتا : يعرف تكامل بيتا (دالة بيتا) من أجل $m>0\ ,\ n>0$ بالعلاقة:

$$B(m,n)=\int_0^1 x^{m-1}(1-x)^{n-1}dx\quad ,\ m>0\ ,\ n>0$$

العلاقة بين دالة غاما ودالة بيتا:

$$B(m,n)=\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}\quad ,\ m>0\ ,\ n>0$$

ملاحظة 15: نظرية النهاية المركزية :

لتكن $X_1,X_2,\ldots\ldots,X_n$ عينة عشوائية لمتغير عشوائي له عزوم من مراتب عالية عندئذ يكون لدينا $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ من النمط الطبيعي المعياري حيث أنّ : $E(X)=\mu\quad ;\quad Var(X)=\sigma^2$

متباينة تشيبتشيف : ليكن X متغيراَ عشوائياَ توقعه μ وتباينه σ^2 فإذا كان a أي عدد موجب فإنّ :

$$P\{|X-\mu|\geq a\}\leq\frac{\sigma^2}{a^2}$$

قانون الأعداد الكبيرة: ليكن $f(x)$ دالة كثافة لـ X بمتوسط μ وتباين محدد σ^2 وليكن $\bar{X}$ المتوسط الحسابي لعينة عشوائية مأخوذة من$f(x)$ حجمها n وليكن ε,δ أي عددين موجبين عندئذ يمكن إيجاد عدد n_0 بحيث أنّ: $P\{|\bar{X}-\mu|<\varepsilon\}\geq 1-\delta$ من اجل جميع قيم $n>n_0$.

الفصل الأول

توزيعات العينة

Sample distributions

إذا كان التوزيع الاحتمالي لمتغير عشوائي X أو التوزيع المشترك لعدة متغيرات عشوائية معروفة فكيف نستطيع استخدام هذه المعرفة للوصول الى شكل التوزيع الاحتمالي لدالة ما في X أو أكثر في مجموعة من المتغيرات العشوائية .

أولاً : إيجاد دالة الكثافة لمتغير عشوائي تابع لمتغير عشوائي أخر دالة كثافته معلومة :

نفرض أنّ دالة الكثافة للمتغير X هي $f(x)$ معلومة وأنّنا نريد معرفة دالة الكثافة للمتغير الجديد Y المعرف بالعلاقة $Y = U(X)$ سوف نعالج ثلاث حالات:

1 – الدالة $U(X)$ متزايدة باطراد :

لدينا في هذه الحالة دالة التوزيع للمتغير Y تكتب بالشكل:

$$G(y) = p(Y \leq y) = p\big(U(X) \leq u(x)\big) = p(X \leq x) = F(x)$$

وباشتقاق الطرفين بالنسبة لـ y نجد:

$$g(y) = \frac{d}{dy} G(y) = \frac{d}{dy} F(x) = \frac{d}{dx} F(x) \frac{dx}{dy} = f(x) \frac{dx}{dy}$$

حيث نعبر عن الطرف الأيمن بدلالة y . أي أنّ دالة الكثافة للمتغير Y تصبح بالشكل:

$$g(y) = f(x) \frac{dx}{dy}\bigg|_{x=u^{-1}(y)}$$

2 – الدالة $U(X)$ متناقصة باطراد:

لدينا في هذه الحالة دالة التوزيع للمتغير Y تكتب بالشكل:

$$G(y) = p(Y \leq y) = p\big(U(X) \leq u(x)\big) = p(X \geq x) = 1 - F(x)$$

وباشتقاق الطرفين بالنسبة لـ y نجد:

$$g(y) = -\frac{d}{dy} G(y) = -\frac{d}{dy} F(x) = \frac{d}{dx} F(x) \left(-\frac{dx}{dy}\right) = f(x) \left(-\frac{dx}{dy}\right)$$

حيث نعبر عن الطرف الأيمن بدلالة y . أي أنّ دالة الكثافة للمتغير Y تصبح بالشكل:

$$g(y) = f(x) \left(-\frac{dx}{dy}\right)\bigg|_{x=u^{-1}(y)}$$

ويمكن دمج الحالتين معاً بالعلاقة الآتية :

$$g(y) = f(x)\left|\frac{dx}{dy}\right|_{x=u^{-1}(y)}$$

مثال: ليكن لدينا متغير عشوائي X دالة كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x) = \begin{cases} 2x & ; \quad 0 \le x \le 1 \\ 0 & ; \quad \text{عدا ذلك} \end{cases}$$

وليكن Y متغير عشوائي آخر معرف بالشكل: $Y = x^3 - 8$.

المطلوب أوجد دالة الكثافة للمتغير Y .

الحل :

طريقة أولى: الدالة المعطاة متزايدة وذلك لأنّ قيم x مو جبة دوما :

$$G(y) = P(Y \le y) = P(X^3 - 8 \le y) = P(X^3 \le y + 8)$$

$$= P\left(X \le \sqrt[3]{y+8}\right) = \int_0^{\sqrt[3]{y+8}} f(x)\,dx$$

$$= \int_0^{\sqrt[3]{y+8}} 2x\,dx = x^2\Big|_0^{\sqrt[3]{y+8}} = (y+8)^{\frac{2}{3}}$$

ومنه نجد دالة الكثافة للمتغير Y باشتقاق دالة التوزيع مباشرة :

$$g(y) = \frac{2}{3}(y+8)^{-\frac{1}{3}} \quad , \quad -8 \le y \le -7$$

طريقة ثانية: بالاعتماد على القانون مباشرة نجد أنّ:

$$g(y) = f(x)\left|\frac{dx}{dy}\right|_{x=u^{-1}(y)} = 2x\frac{1}{3}(y+8)^{-\frac{2}{3}}\Big|_{x=(y+8)^{\frac{1}{3}}}$$

$$= \frac{2}{3}(y+8)^{\frac{1}{3}}(y+8)^{-\frac{2}{3}} = \frac{2}{3}(y+8)^{-\frac{1}{3}}$$

أي أن:

$$g(y) = \frac{2}{3}(y+8)^{-\frac{1}{3}} \quad , \quad -8 \le y \le -7$$

وهي نفس النتيجة السابقة التي حصلنا عليها .

3 – الدالة $U(X)$ ليست دالة مطردة فهي متزايدة تارة ومتناقصة تارة أخرى:

ففي هذه الحالة نقسم مجال X إلى مجالات تكون فيها الدالة $U(X)$ في كل منها وحيدة الطور أي متزايدة باطراد ثم نطبق القاعدة السابقة في كل من هذه المجالات.

لنأخذ حالة تكون فيها الدالة تمر بطورين فقط , متزايدة ثم متناقصة فنجد:

$$g(y)=f(x_1)\frac{dx_1}{dy}\Big|_{x_1=u_1^{-1}(y)}+f(x_2)\left(-\frac{dx_2}{dy}\right)\Big|_{x_2=u_2^{-1}(y)}$$

$$=f(x_1)\left|\frac{dx_1}{dy}\right|\Bigg|_{x_1=u_1^{-1}(y)}+f(x_2)\left|\frac{dx_2}{dy}\right|\Bigg|_{x_2=u_2^{-1}(y)}$$

ويمكن تعميم النتيجة إلى الحالة التي يقابل فيها كل قيمة لـ y عدد من القيم لـ x ولتكن n مثلاً فنكتب :

$$g(y)=\sum_{i=1}^{n}f(x_i)\left|\frac{dx_i}{dy}\right|\Bigg|_{x_i=u_i^{-1}(y)}$$

مثال: ليكن لدينا متغير عشوائي X له التوزيع الطبيعي المعياري اي أنّ دالة كثافته معرفة بالشكل التالي :

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ ;x\in R$$

وليكن $Y=X^2$ أوجد دالة الكثافة الاحتمالية للمتغير Y .

الحل:

بما أنّ: $y=x^2\Rightarrow x=\pm\sqrt{y}$

فإنّنا نجد قيمتان لـ x هما : $x_1=\sqrt{y}\ ,\ \ x_2=-\sqrt{y}$

ومنه نجد أنّ:

$$g(y)=f(x_1)\left|\frac{dx_1}{dy}\right|+f(x_2)\left|\frac{dx_2}{dy}\right|$$

$$=\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{x_1^2}{2}}+e^{-\frac{x_2^2}{2}}\right)\left(\frac{1}{2\sqrt{y}}\right)$$

نعوض الآن كل من x_1,x_2 بما يساويها فنحصل على:

$$g(y)=\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{y}{2}}+e^{-\frac{y}{2}}\right)\left(\frac{1}{2\sqrt{y}}\right)$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}y^{-\frac{1}{2}}=\frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}y^{-\frac{1}{2}}e^{-\frac{y}{2}}$$

أي أن:

$$g(y)=\frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}y^{-\frac{1}{2}}e^{-\frac{y}{2}}\quad ,\ y\geq 0$$

وهي دالة الكثافة لتوزيع كاي مربع درجة حريته الواحد.

ثانياً: إيجاد دالة الكثافة الاحتمالية المشتركة لمجموعة من المتغيرات العشوائية التابعة لمجموعة أخرى من المتغيرات العشوائية دالة كثافتها الاحتمالية معلومة :

لتكن لدينا X_1, X_2, X_3 ثلاثة متغيرات عشوائية كثافتها المشتركة معلومة , ولنفرض أنّ Y_2, Y_2, Y_3 ثلاثة متغيرات عشوائية أخرى معرفة بالشكل :

$$Y_1 = U_1(X_1, X_2, X_3)$$
$$Y_2 = U_2(X_1, X_2, X_3)$$
$$Y_3 = U_3(X_1, X_2, X_3)$$

المطلوب إيجاد الكثافة المشتركة للمتغيراتY_2, Y_2, Y_3.

الحل:

إنّ دالة الكثافة المشتركة لـ Y_1, Y_2, Y_3 تعطى بالعلاقة :

$$g(y_1, y_2, y_3) = f(x_1, x_2, x_3)|J|$$

حيث أنّ J هو معين اليعقوبي ويعطى كما يلي :

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix}$$

مثال: لتكن X, Y عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي له التوزيع الطبيعي المعياري , وبفرض أنّ U, V متغيرين عشوائيين مستقلين جديدين معرفين بالعلاقتين التاليتين:

$$U = X + Y \; ; \; V = X - Y$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذين المتغيرين العشوائيين U, V .

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من هذين المتغيرين العشوائيين U, V .

3 – أثبت أنّ هذين المتغيرين العشوائيين U, V مستقلان عشوائياً .

الحل:

1 – بما أنّ X, Y متغيران عشوائيان مستقلان فإنّ دالة الكثافة المشتركة تعطى بالعلاقة:

$$f(x, y) = f(x)f(y) = (\frac{1}{\sqrt{2\pi}})^2 e^{-(\frac{x^2+y^2}{2})}$$

بفرض أنّ:

$$\left.\begin{matrix} u = x + y \\ v = x - y \end{matrix}\right\} \Rightarrow \left\{\begin{matrix} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{matrix}\right.$$

ثم نحسب معين اليعقوبي بالشكل التالي:

$$J=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{vmatrix}=-\frac{1}{2}\Longrightarrow |J|=\frac{1}{2}$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة لـ U,V تعطى بالشكل:

$$g(u,v)=f(x,y)|J|=\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right)\left|-\frac{1}{2}\right|\Bigg|_{\substack{x=\frac{u+v}{2}\\ y=\frac{u-v}{2}}}$$

$$=\frac{1}{4\pi}e^{-\frac{1}{4}(u^2+v^2)}=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}$$

أي أنّ :

$$g(u,v)=\frac{1}{4\pi}e^{-\frac{1}{4}(u^2+v^2)}\quad ;\ -\infty< u,v<\infty$$

2 – لنوجد دالة الكثافة الهامشية لكل من المتغيرين U,V :

$$g(u)=\int_{-\infty}^{\infty}g(u,v)dv=\int_{-\infty}^{\infty}\frac{1}{4\pi}e^{-\frac{1}{4}(u^2+v^2)}dv$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}}_{N(0,2)}dv=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}$$

أي أن:

$$g(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}\quad ;\ -\infty<u<\infty$$

$$g(v)=\int_{-\infty}^{\infty}g(u,v)du=\int_{-\infty}^{\infty}\frac{1}{4\pi}e^{-\frac{1}{4}(u^2+v^2)}du$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}}_{N(0,2)}du=\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}$$

أي أن:

$$g(v)=\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}\quad ;\ -\infty<v<\infty$$

3 – نلاحظ أن:

$$g(u,v)=\frac{1}{4\pi}e^{-\frac{1}{4}(u^2+v^2)}$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{4}}\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{4}}=g(u)g(v)$$

وبالتالي فإنّ هذين المتغيرين مستقلين وتوزيع كل منهما هو التوزيع الطبيعي وسيطاه $\mu = 0$, $\sigma^2 = 2$ أي أنّ توزيع كل منهما هو التوزيع الطبيعي $N(0,2)$.

مثال: ليكن X متغير عشوائي له التوزيع الطبيعي المعياري وليكن Z متغيرا عشوائيّا أخراً له توزيع كاي مربّع درجة حريته n وبفرض أنّ X,Z مستقلان, وأن T,Y متغيرين عشوائيين معرفين بالعلاقتين التاليتين:

$$T = \frac{X}{\sqrt{\frac{Z}{n}}} \quad , \quad Y = Z$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين T,Y .

2 – أوجد دالة الكثافة الاحتمالية للمتغير T .

الحل:

1 – بما أنّ: $Z \sim \chi^2(n)$; $X \sim N(0,1)$ وأنّ X,Z مستقلان , فإن:

$$f(x,z) = f(x)f(z) = \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right]\left[\frac{1}{(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(z)^{\frac{n}{2}-1} e^{-\frac{z}{2}}\right]$$

نفرض أن:

$$\left.\begin{array}{l} t = \dfrac{x}{\sqrt{\frac{z}{n}}} \\ y = z \end{array}\right\} \Rightarrow \left\{\begin{array}{l} x = t\sqrt{\dfrac{y}{n}} \\ z = y \end{array}\right.$$

ثم نحسب معين اليعقوبي بالشكل التالي:

$$\mathrm{J} = \begin{vmatrix} \frac{\partial \mathrm{x}}{\partial \mathrm{t}} & \frac{\partial \mathrm{x}}{\partial \mathrm{y}} \\ \frac{\partial \mathrm{z}}{\partial \mathrm{t}} & \frac{\partial \mathrm{z}}{\partial \mathrm{y}} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{y}{n}} & \text{مقدارما} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y}{n}}$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة لـ T,Y تعطى بالشكل:

$$g(t,y) = f(x,z)|J|\Big|_{\substack{x=t\sqrt{\frac{y}{n}} \\ z=y}}$$

$$= \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right]\left[\frac{1}{(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(z)^{\frac{n}{2}-1} e^{-\frac{z}{2}}\right]\sqrt{\frac{y}{n}}\Big|_{\substack{x=t\sqrt{\frac{y}{n}} \\ z=y}}$$

$$= \frac{1}{\sqrt{2\pi}(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(y)^{\frac{n}{2}-1} e^{-\frac{t^2 y}{2n}} e^{-\frac{y}{2}}\sqrt{\frac{y}{n}}$$

إذا أصبح لدينا:

$$g(t,y)=\frac{1}{\sqrt{n\pi}\ (2)^{\frac{n+1}{2}}\ \Gamma\left(\frac{n}{2}\right)}e^{-\frac{y}{2}(1+\frac{t^2}{n})}y^{\frac{n-1}{2}}\quad ;\quad y>0\,;t\in R$$

2 – وبالتالي فإنّ دالة الكثافة الهامشية لـ T تعطى بالشكل:

$$f(t)=\int_0^\infty g(t,y)dy=\int_0^\infty \frac{1}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{y}{2}(1+\frac{t^2}{n})}y^{\frac{n-1}{2}}dy$$

نجري تغيير في المتحول بالشكل:

$$z=\frac{y}{2}\left(1+\frac{t^2}{n}\right)\Longrightarrow y=\frac{2z}{(1+\frac{t^2}{n})}\Longrightarrow dy=\frac{2dz}{(1+\frac{t^2}{n})}$$

ومنه نجد :

$$f(t)=a\int_0^\infty e^{-z}\left(\frac{2z}{1+\frac{t^2}{n}}\right)\frac{2dz}{(1+\frac{t^2}{n})}=\frac{a2^{\frac{n+1}{2}}}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}}\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz$$

علما أنّ:
$$a=\frac{1}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)}$$

وبالتالي يمكن أنّ نكتب:

$$f(t)=\frac{2^{\frac{n+1}{2}}}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)(1+\frac{t^2}{n})^{\frac{n+1}{2}}}\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz$$

ومن دالة غاما نجد:
$$\Gamma(\vartheta)=\int_0^\infty x^{\vartheta-1}e^{-x}dx\Longrightarrow\frac{n-1}{2}=\vartheta-1\Longrightarrow\vartheta=\frac{n+1}{2}$$

أي أنّ:
$$\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz=\Gamma\left(\frac{n+1}{2}\right)$$

نعوض مباشرةً فنجد أنّ :

$$f(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)(1+\frac{t^2}{n})^{\frac{n+1}{2}}}=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

وهي دالة الكثافة الاحتمالية لتوزيع ستيودنت درجة حريته n .

ويمكن صياغة المبرهنة بشكل أخر : إذا كان لدينا $Y\sim\chi^2_{(n)}$, $X\sim N(0,1)$ وحيث أنّ X,Y مستقلان فهذا يؤدي إلى أنّ $T=\frac{X}{\sqrt{\frac{y}{n}}}$ له توزيع ستيودنت درجة حريّته n .

مثال: ليكن U,V متغيرين عشوائيين مستقلين لكل منهما توزيع كاي مربع درجة حريتهما على الترتيب r_1,r_2

وبفرض أنّ F,Z متغيرين عشوائيين معرفين بالعلاقتين التاليتين:

$$F=\frac{U/r_1}{V/r_2}\quad ,\quad Z=V$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين F,Z .

2 – أوجد دالة الكثافة الاحتمالية للمتغير F ثم بين نوعه .

الحل:

1 – لدينا

$$U\sim\chi^2(r_1) \Rightarrow f(u) = \frac{1}{2^{\frac{r_1}{2}}}\frac{1}{\Gamma(\frac{r_1}{2})}u^{\frac{r_1}{2}-1}e^{-\frac{u}{2}}$$
$$V\sim\chi^2(r_2) \Rightarrow f(v) = \frac{1}{2^{\frac{r_2}{2}}\Gamma(\frac{r_2}{2})}v^{\frac{r_2}{2}-1}e^{-\frac{v}{2}}$$

وبما أنّ U,V متغيرين عشوائيين مستقلين, فإن :

$$f(u,v) = \frac{1}{(2)^{\frac{r_1+r_2}{2}}\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})}u^{\frac{r_1}{2}-1}v^{\frac{r_2}{2}-1}e^{-\frac{u+v}{2}}$$

نفرض أن:

$$\left.\begin{matrix} f == \dfrac{u/r_1}{v/r_2} \\ z = v \end{matrix}\right\} \Longrightarrow \begin{matrix} u = \dfrac{r_1}{r_2}fz \\ v = z \end{matrix}$$

ثم نحسب معين اليعقوبي بالشكل التالي:

$$J = \frac{\partial(u,v)}{\partial(f,z)} = \begin{vmatrix} \frac{\partial u}{\partial f} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial f} & \frac{\partial v}{\partial z} \end{vmatrix} = \begin{vmatrix} \frac{r_1}{r_2}z & \frac{r_1}{r_2}f \\ 0 & 1 \end{vmatrix} = \frac{r_1}{r_2}z$$

وبالتالي فإنّ دالة الكثافة الاحتمالية المشتركة تعطى بالشكل:

$$g(f,z) = f(u,v)|J|$$
$$= \left.\frac{1}{(2)^{\frac{r_1+r_2}{2}}\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})}u^{\frac{r_1}{2}-1}v^{\frac{r_2}{2}-1}e^{-(\frac{u+v}{2})}\left|\frac{r_1}{r_2}z\right|\right|_{\substack{u=\frac{r_1}{r_2}fz \\ v=z}}$$
$$= \frac{1}{(2)^{\frac{r_1+r_2}{2}}\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}\left(\frac{r_1}{r_2}f.z\right)^{\frac{r_1}{2}-1}z^{\frac{r_2}{2}-1}e^{-\frac{z}{2}\left(\frac{r_1}{r_2}f+1\right)}\frac{r_1}{r_2}z$$
$$= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}f^{\frac{r_1}{2}-1}}{(2)^{\frac{r_1+r_2}{2}}\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}z^{\frac{r_1+r_2}{2}-1}e^{-\frac{z}{2}\left(\frac{r_1}{r_2}f+1\right)} \quad , \; 0 < f,z < \infty$$

2 – لنوجد الآن دالة الكثافة الهامشية لـ f بالشكل :

$$g(f) = \int_{-\infty}^{\infty} g(f,z)dz$$

$$= \int_0^{\infty} \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1}}{(2)^{\frac{r_1+r_2}{2}}\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} z^{\frac{r_1+r_2}{2}-1} e^{-\frac{1}{2}\left(\frac{r_1}{r_2}f+1\right)z} dz$$

لإجراء التكامل نفرض أنّ: $y = \frac{1}{2}\left(\frac{r_1}{r_2} f + 1\right) z \Rightarrow z = \frac{2y}{\frac{r_1}{r_2}f+1} \Rightarrow dz = \frac{2dy}{\frac{r_1}{r_2}f+1}$

$$g(f) = \int_0^{\infty} \left(\frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1} \left(\frac{2y}{\frac{r_1}{r_2}f+1}\right)^{\frac{r_1+r_2}{2}-1}}{(2)^{\frac{r_1+r_2}{2}}\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} e^{-y} \frac{2}{\frac{r_1}{r_2}f+1} \right) dy$$

$$= \left(\frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1} \left(1+\frac{r_1}{r_2}f\right)^{-\frac{r_1+r_2}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \right) \underbrace{\int_0^{\infty} (y)^{\frac{r_1+r_2}{2}-1} e^{-y} dy}_{\Gamma\left(\frac{r_1+r_2}{2}\right)}$$

$$= \left(\frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1} \left(1+\frac{r_1}{r_2}f\right)^{-\frac{r_1+r_2}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \right) \Gamma\left(\frac{r_1+r_2}{2}\right) \quad , \quad 0 < f < \infty$$

أي أن:

$$g(f) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1} \left(1+\frac{r_1}{r_2}f\right)^{-\frac{r_1+r_2}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \quad , \quad 0 < f < \infty$$

وهي دالة احتمالية لتوزيع فيشر.

مثال: لتكن X,Y عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي دالة كثافه معرفة بالشكل:

$$f(x) = e^{-x} \;\; ; \;\; x \geq 0$$

وبفرض أنّ U,V متغيران عشوائيان معرفان بالشكل:

$$U = X + Y \quad , \quad V = \frac{X}{Y}$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذين المتغيرين العشوائيين U,V .

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من هذين المتغيرين العشوائيين U,V .

3 – أثبت أنّ هذين المتغيرين العشوائيين U,V مستقلان عشوائياً .

الحل:

1 – بما أن X,Y متغيرين عشوائيين مستقلين, فإن :

$$f(x,y)=f(x)f(y)=e^{-(x+y)}$$

بفرض أنّ:

$$\left.\begin{matrix} u=x+y \\ v=\frac{x}{y} \end{matrix}\right\} \Longrightarrow \left\{\begin{matrix} x=\frac{uv}{v+1} \\ y=\frac{u}{v+1} \end{matrix}\right.$$

ثم نحسب معين اليعقوبي بالشكل التالي:

$$J=\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}=\begin{vmatrix} \frac{v}{v+1} & \frac{u}{(v+1)^2} \\ \frac{1}{v+1} & -\frac{u}{(v+1)^2} \end{vmatrix}=-\frac{u}{(v+1)^2}\Longrightarrow |J|=\frac{u}{(v+1)^2}$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة لـ U,V تعطى بالشكل:

$$g(u,v)=f(x,y)|J|=e^{-(x+y)}\left|-\frac{u}{(v+1)^2}\right|\Bigg|_{\substack{x=\frac{uv}{v+1} \\ y=\frac{u}{v+1}}}$$

$$=e^{-u}\frac{u}{(v+1)^2}=\frac{ue^{-u}}{(v+1)^2}$$

أي أنّ :

$$g(u,v)=\frac{ue^{-u}}{(v+1)^2} \quad ; \quad u,v>0$$

2 – لنوجد دالة الكثافة الهامشية لكل من المتغيرين U,V :

$$g(u)=\int_{-\infty}^{\infty}g(u,v)dv=\int_0^{\infty}\frac{ue^{-u}}{(v+1)^2}dv$$

$$=ue^{-u}\int_0^{\infty}\frac{dv}{(v+1)^2}=ue^{-u}$$

أي أنّ :

$$g(u)=ue^{-u} \quad ; \quad u>0$$

$$g(v)=\int_{-\infty}^{\infty}g(u,v)du=\int_0^{\infty}\frac{ue^{-u}}{(v+1)^2}du$$

$$=\frac{1}{(v+1)^2}\int_0^{\infty}ue^{-u}du=\frac{1}{(v+1)^2}$$

أي أنّ :

$$g(v)=\frac{1}{(v+1)^2} \quad ; \quad v>0$$

3 – نلاحظ أن:

$$g(u,v)=\frac{ue^{-u}}{(v+1)^2}=(ue^{-u})\left(\frac{1}{(v+1)^2}\right)=g(u)g(v)$$

وبالتالي فإنّ هذين المتغيرين مستقلين .

مثال: لتكن X, Y, Z ثلاثة متغيرات عشوائية دالة كثافتهما الاحتمالية المشتركة معرفة بالشكل:

$$f(x,y,z)=\frac{6}{(1+x+y+z)^4} \quad , \qquad x>0\ ,\ y>0\ ,\ z>0$$

وبفرض أنّ U, V, W ثلاثة متغيرات عشوائية أخرى معرفة بالشكل:

$$U=X+Y+Z \quad , \quad V=Y+Z \quad , \quad W=Z$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذه المتغيرات العشوائية أنّ U, V, W .

2 – أوجد دالة الكثافة الاحتمالية الهامشية للمتغير $U=X+Y+Z$.

الحل:

1 – نفرض أن:

$$\left.\begin{matrix} u=x+y+z \\ v=y+z \\ w=z \end{matrix}\right\} \Rightarrow \begin{cases} x=u-v \\ y=v-w \\ z=w \end{cases}$$

ثم نحسب معين اليعقوبي بالشكل التالي:

$$J=\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}=\begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{vmatrix}=1$$

وبالتالي فإنّ دالة الكثافة الاحتمالية المشتركة تعطى بالشكل:

$$f(u,v,w)=f(x,y,z)|J|\ \Big|_{\substack{x=u-v \\ y=v-w \\ z=w}}=\frac{6}{(1+u)^4}$$

أي أنّ دالة الكثافة الاحتمالية المشتركة لـ U, V, W تصبح بالشكل: :

$$f(u,v,w)=\frac{6}{(1+u)^4} \qquad ; \qquad 0<w<v<u<\infty$$

2 – لنوجد الآن دالة الكثافة الهامشية $f(u)$ بالشكل :

$$f(u)=\int_0^u\int_0^v \frac{6}{(1+u)^4}\,dwdv=\frac{3u^2}{(1+u)^4} \quad , \ u\geq 0$$

مثال: لتكن X_1, X_2, X_3 عينة عشوائية حجمها $n = 3$ مأخوذة من مجتمع إحصائي دالة كثافه معرفة بالشكل:

$$f_X(x) = e^{-x} \qquad ; x > 0$$

وبفرض أنّ Y_1 , Y_2 , Y_3 ثلاثة متغيرات عشوائية معرفة بالشكل:

$$Y_1 = \frac{X_1}{X_1+X_2} , Y_2 = \frac{X_1+X_2}{X_1+X_2+X_3} , Y_3 = X_1 + X_2 + X_3$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذه المتغيرات العشوائية Y_1 , Y_2 , Y_3.

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من هذه المتغيرات العشوائية Y_1 , Y_2 , Y_3 .

3 – أثبت أنّ هذه المتغيرات العشوائية Y_1 , Y_2 , Y_3 مستقلة عشوائياً .

الحل:

بما أنّ المتغيرات العشوائية X_1, X_2, X_3 مستقلة عشوائياً فإنّ دالة الكثافة المشتركة تعطى بالشكل:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) \overset{\text{من الاستقلال}}{=} f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3)$$
$$= e^{-x_1} e^{-x_2} e^{-x_3} = e^{-\sum_{i=1}^{3} x_i}$$

وبفرض أن :

$$\left.\begin{aligned} y_1 &= \frac{x_1}{x_1+x_2} \\ y_2 &= \frac{x_1+x_2}{x_1+x_2+x_3} \\ y_3 &= x_1 + x_2 + x_3 \end{aligned}\right\} \Longrightarrow \begin{cases} x_1 = y_1 y_2 y_3 \\ x_2 = (1-y_1) y_2 y_3 \\ x_3 = (1-y_2) y_3 \end{cases}$$

لنوجد الآن محدد اليعقوبي:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix} = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ -y_2 y_3 & (1-y_1) y_3 & (1-y_1) y_2 \\ 0 & -y_3 & (1-y_2) \end{vmatrix}$$

$$= y_2 y_3 (1-y_1) y_3 (1-y_2) + y_1 y_2 (-y_2 y_3)(-y_3) + y_2 y_3 (1-y_1) y_2 y_3 +$$
$$+ (y_1 y_3)(y_2 y_3)(1-y_2) = y_2 y_3^2 - y_1 y_2 y_3^2 - y_2^2 y_3^2 + y_1 y_2^2 y_3^2 + y_1 y_2^2 y_3^2 +$$
$$y_2^2 y_3^2 - y_1 y_2^2 y_3^2 + y_1 y_2 y_3^2 - y_1 y_2^2 y_3^2 = y_2 y_3^2$$

وبالتالي فإنّ دالة الكثافة المشتركة للمتغيرات العشوائية Y_1 , Y_2 , Y_3 تعطى بالشكل:

$$g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = f_{X_1,X_2,X_3}(x_1,x_2,x_3)\, |J| \Big|_{\substack{x_1 = y_1 y_2 y_3 \\ x_2 = (1-y_1) y_2 y_3 \\ x_3 = (1-y_2) y_3}}$$

بالتعويض في العلاقة السابقة نجد:

$$g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = e^{-y_3}y_2y_3^2 = y_2y_3^2e^{-y_3} = 1\times 2y_2 \times \frac{1}{2}y_3^2e^{-y_3} \quad ;\begin{cases} y_3 \geq 0 \\ 0\leq y_2 \leq 1 \\ 0\leq y_1\leq 1\end{cases}$$

$$= [(1)]\left[\frac{\Gamma(2+1)}{\Gamma(2)\Gamma(1)}y_2^{2-1}(1-y_2)^{1-1}\right]\left[\frac{1^3}{\Gamma(3)}y_3^{3-1}e^{-y_3}\right]$$

$$= \underbrace{g_{Y_1}(y_1)}_{\text{منتظم على}[0,1]} \quad \underbrace{g_{Y_2}(y_2)}_{\text{بتاوي وسيطاه}\, m=2,n=1} \quad \underbrace{g_{Y_3}(y_3)}_{\text{غمّاوي وسيطاه}\,\lambda=3,\alpha=1} \quad ; \; 0\leq y_1,y_2\leq 1, y_3\geq 0$$

$\Leftarrow$ Y_1 ,Y_2 ,Y_3 مستقلة عشوائياً.

أو يمكننا اثبات الاستقلال العشوائي بالشكل التقليدي التالي:

$$g_{Y_1}(y_1) = \int_0^1\left(\int_0^\infty y_2y_3^2e^{-y_3}dy_3\right)dy_2 = \int_0^1 y_2\Gamma(3)dy_2 = \int_0^1 2y_2dy_2 = 1,\; 0\leq y_1\leq 1$$

$$g_{Y_2}(y_2) = \int_0^1\left(\int_0^\infty y_2y_3^2e^{-y_3}dy_3\right)dy_1 = \int_0^{y_2} 2y_2dy_1 = 2y_2\,,\; 0\leq y_2\leq 1$$

$$g_{Y_3}(y_3) = \int_0^1\left(\int_0^1 y_2y_3^2e^{-y_3}dy_2\right)dy_1 = y_3^2e^{-y_3}\int_0^1\left(\int_0^1 y_2dy_2\right)dy_1$$

$$= e^{-y_3}y_3^2\int_0^1\frac{1}{2}dy_1 = \frac{1}{2}y_3^2e^{-y_3}\,, y_3\geq 0$$

نلاحظ أنّ:

$$g_{Y_1}(y_1)\, g_{Y_2}(y_2)\, g_{Y_3}(y_3) = 1\times 2y_2\times\frac{1}{2}y_3^2e^{-y_3} = y_2y_3^2e^{-y_3}$$

$$= g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)$$

وبالتالي فإنّ المتغيرات العشوائية Y_1 ,Y_2 ,Y_3 مستقلة عشوائياً.

مثال: لتكن X_1,X_2,X_3 عينة عشوائية حجمها $n = 3$ مأخوذة من مجتمع إحصائي دالة كثافه معرفة بالشكل:

$$f_X(x) = e^{-x} \qquad ;x>0$$

وبفرض أنّ Y_1 ,Y_2 ,Y_3 ثلاثة متغيرات عشوائية معرفة بالشكل:

$$Y_1 = \frac{X_1}{X_1+X_2+X_3}\,, Y_2 = \frac{X_1+X_2}{X_1+X_2+X_3}\,, Y_3 = X_1+X_2+X_3$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذه المتغيرات العشوائية Y_1 ,Y_2 ,Y_3.

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من هذه المتغيرات العشوائية Y_1 ,Y_2 ,Y_3 .

3 – أثبت فيما إذا كانت هذه المتغيرات العشوائية Y_1 ,Y_2 ,Y_3 مستقلة عشوائياً أم لا .

الحل:

بما أنّ المتغيرات العشوائية X_1,X_2,X_3 مستقلة عشوائياً فإنّ دالة الكثافة المشتركة تعطى بالشكل:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) \overset{\text{من الاستقلال}}{=} f_{X_1}(x_1)\,.f_{X_2}(x_2)\,.f_{X_3}(x_3)$$

$$= e^{-x_1}e^{-x_2}e^{-x_3} = e^{-\sum_{i=1}^{3} x_i}$$

وبفرض أن :

$$\left.\begin{array}{r} y_1 = \frac{x_1}{x_1+x_2+x_3} \\ y_2 = \frac{x_1+x_2}{x_1+x_2+x_3} \\ y_3 = x_1+x_2+x_3 \end{array}\right\} \Rightarrow \left\{\begin{array}{c} x_1 = y_1y_3 \\ x_2 = (y_2-y_1)y_3 \\ x_3 = (1-y_2)y_3 \end{array}\right.$$

لنوجد الآن محدد اليعقوبي:

$$\mathrm{J} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix} = \begin{vmatrix} y_3 & 0 & y_1 \\ -y_3 & y_3 & (y_2-y_1) \\ 0 & -y_3 & (1-y_2) \end{vmatrix}$$

$$= y_3[y_3(y_2-y_1)+y_1y_3]+(1-y_2)y_3^2 = y_2y_3^2-y_1y_3^2+y_1y_3^2+y_3^2-y_2^2y_3^2 = y_3^2$$

وبالتالي فإنّ دالة الكثافة المشتركة للمتغيرات العشوائية Y_1 ,Y_2 ,Y_3 تعطى بالشكل:

$$g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = f_{X_1,X_2,X_3}(x_1,x_2,x_3)\ |\mathrm{J}| \ \ \begin{array}{l} {\scriptstyle x_1=y_1y_3} \\ {\scriptstyle x_2=(y_2-y_1)y_3} \\ {\scriptstyle x_3=(1-y_2)y_3} \end{array}$$

بالتعويض في العلاقة السابقة نجد دالة الكثافة المشتركة:

$$g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = y_3^2e^{-y_3} \ \ ; \ \ 0 \le y_1 \le y_2 \le 1, y_3 \ge 0$$

لنوجد دوال الكثافات الهامشية بالشكل:

$$g_{Y_1}(y_1) = \int_0^1\left(\int_0^\infty y_3^2e^{-y_3}dy_3\right)dy_2 = \int_0^1 \Gamma(3)dy_2 = \int_0^1 2dy_2 = 2 \ , \ 0 \le y_1 \le \frac{1}{2}$$

$$g_{Y_2}(y_2) = \int_0^{y_2}\left(\int_0^\infty y_3^2e^{-y_3}dy_3\right)dy_1 = \int_0^{y_2} \Gamma(3)dy_1 = \int_0^{y_2} 2dy_1 = 2y_2 \ , \ 0 \le y_2 \le 1$$

$$g_{Y_3}(y_3) = \int_0^{\frac{1}{2}}\left(\int_0^1 y_3^2e^{-y_3}dy_2\right)dy_1 = y_3^2e^{-y_3}\int_0^{\frac{1}{2}}\left(\int_0^1 dy_2\right)dy_1$$

$$= y_3^2e^{-y_3}\int_0^{\frac{1}{2}} dy_1 = \frac{1}{2}y_3^2e^{-y_3} \ , y_3 \ge 0$$

نلاحظ أنّ:

$$g_{Y_1}(y_1)\ g_{Y_2}(y_2)\ g_{Y_3}(y_3) = 2 \times 2y_2 \times \frac{1}{2}y_3^2e^{-y_3} = 2y_2y_3^2e^{-y_3}$$

$$\neq e^{-y_3}y_3^2 = g_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)$$

وبالتالي فإنّ المتغيرات العشوائية Y_1 ,Y_2 ,Y_3 غير مستقلة عشوائياً.

مثال: ليكن X_1,X_2 متغيرين عشوائيين غمّاويين مستقلين , وسيطاهما α,β على الترتيب , دالة كثافتهما الاحتمالية المشتركة معرفة بالشكل التالي:

$$f(x_1,x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)}x_1^{\alpha-1}x_2^{\beta-1}e^{-x_1-x_2} \ \ ; \ \ 0 < x_1,x_2 < \infty \ \ \alpha,\beta > 0$$

وبفرض أنّ Y_1 , Y_2 متغيرين عشوائيين معرفين بالشكل:

$$Y_1 = X_1 + X_2 \quad , \quad Y_2 = \frac{X_1}{X_1+X_2}$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين Y_1 , Y_2.

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من المتغيرين العشوائيين Y_1 , Y_2 .

الحل:

1 – لدينا من الفرض أنّ:

$$\left.\begin{array}{r} y_1 = x_1 + x_2 \\ y_2 = \frac{x_1}{x_1+x_2} \end{array}\right\} \Rightarrow \left\{\begin{array}{c} x_1 = y_1 y_2 \\ x_2 = y_1(1-y_2) \end{array}\right.$$

لنوجد الآن معين اليعقوبي J :

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix} = -y_1 \Rightarrow |J| = |-y_1| = y_1$$

بالتعويض نجد أنّ:

$$\begin{aligned} g(y_1,y_2) &= f(x_1,x_2)|J| \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}(y_1 y_2)^{\alpha-1}(y_1 - y_1 y_2)^{\beta-1} e^{-y_1} y_1 \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}(y_1)^{\alpha+\beta-1} e^{-y_1}(y_2)^{\alpha-1}(1-y_2)^{\beta-1} \end{aligned}$$

وبالتالي فإنّ دالة الكثافة المشتركة للمتغيرين العشوائيين Y_1 , Y_2 تعطى بالشكل:

$$g(y_1,y_2) = \frac{{y_1}^{\alpha+\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} e^{-y_1}(y_2)^{\alpha-1}(1-y_2)^{\beta-1} \ ; \ y_1 > 0 \ , \ 0 < y_2 < 1 \ , \ \alpha,\beta > 0$$

2 – إنّ دالة الكثافة الهامشية للمتغير العشوائي Y_1 تعطى بالشكل:

$$\begin{aligned} g(y_1) &= \int_0^\infty g(y_1,y_2) dy_2 \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} {y_1}^{\alpha+\beta-1} e^{-y_1} \int_0^1 (y_2)^{\alpha-1}(1-y_2)^{\beta-1}\, dy_2 \\ &= \frac{{y_1}^{\alpha+\beta-1} e^{-y_1}}{\Gamma(\alpha)\Gamma(\beta)} Be(\alpha,\beta) \\ &= \frac{{y_1}^{\alpha+\beta-1} e^{-y_1}}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{{y_1}^{\alpha+\beta-1} e^{-y_1}}{\Gamma(\alpha+\beta)} \end{aligned}$$

أي أنّ:

$$g(y_1) = \frac{1}{\Gamma(\alpha+\beta)} {y_1}^{\alpha+\beta-1} e^{-y_1} \quad ; \ y_1 > 0 \ , \quad \alpha, \beta > 0$$

أنّ دالة الكثافة الهامشية للمتغير العشوائي Y_2 تعطى بالشكل:

$$\begin{aligned} g(y_2) &= \int_0^1 g(y_1, y_2) dy_1 \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (y_2)^{\alpha-1} (1-y_2)^{\beta-1} \int_0^\infty {y_1}^{\alpha+\beta-1} e^{-y_1} \, dy_1 \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (y_2)^{\alpha-1} (1-y_2)^{\beta-1} \end{aligned}$$

أي أنّ:

$$g(y_2) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (y_2)^{\alpha-1} (1-y_2)^{\beta-1} \ ; \quad 0 < y_2 < 1 \, , \quad \alpha, \beta > 0$$

مثال: لتكن X, Y عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي يخضع لتوزيع كاي مربع درجة حريته (2) , وبفرض أنّ U, V متغيران عشوائيان معرفان بالشكل:

$$U = \frac{1}{2}(X - Y) \quad , \quad V = Y$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذين المتغيرين العشوائيين U, V .

2 – أوجد دالة الكثافة الاحتمالية الهامشية للمتغير العشوائي U .

الحل:

1 – بما أن X, Y متغيرين عشوائيين مستقلين, فإن :

$$\begin{aligned} f(x,y) &= f(x) f(y) \\ &= \left[\frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{1}{2}x} \right] \left[\frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y} \right] \\ &= \left[\frac{1}{2} e^{-\frac{1}{2}x} \right] \left[\frac{1}{2} e^{-\frac{1}{2}y} \right] \\ &= \frac{1}{4} e^{-\left(\frac{x+y}{2}\right)} \ ; \quad x, y > 0 \end{aligned}$$

وبفرض أنّ:

$$\left. \begin{aligned} u &= \tfrac{1}{2}(x-y) \\ v &= y \end{aligned} \right\} \Longrightarrow \left\{ \begin{aligned} x &= 2u + v \\ y &= v \end{aligned} \right.$$

و أنّ هذا التحويل هو تحويل مطابق من $\mathcal{A} = \{(x,y); 0 < x < \infty \, , 0 < y < \infty \ \}$ إلى

$$\mathcal{B} = \{(u,v)\, ; -2u < v\ \ ,0 < v, -\infty < u < \infty\}$$

عندئذ فإن محدد اليعقوبي يعطى بالشكل التالي:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2 \Rightarrow |J| = 2$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة لـ U,V تعطى بالشكل:

$$g(u,v) = f(x,y)|J| = \frac{1}{4}e^{-\left(\frac{x+y}{2}\right)}|2|\Big|_{\substack{x=2u+v \\ y=v}}$$

$$= \frac{1}{2}e^{-u}e^{-v}$$

أي أنّ :

$$g(u,v) = \frac{1}{2}e^{-u-v} \quad ; \quad (u,v) \in \mathcal{B}$$

2 – لنوجد دالة الكثافة الهامشية للمتغير العشوائي U :

$$g(u) = \begin{cases} \int_{-2u}^{\infty} \left(\frac{1}{2}e^{-u-v}\right) dv = \frac{1}{2}e^{u} & ; -\infty < u < 0 \\ \int_{0}^{\infty} \left(\frac{1}{2}e^{-u-v}\right) dv = \frac{1}{2}e^{-u} & ; \ 0 \le u < \infty \end{cases}$$

أي أنّ :

$$g(u) = \frac{1}{2}e^{-|u|} \quad ; \ -\infty < \ u < \infty$$

مثال: لتكن X,Y عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي يخضع لتوزيع كاي مربع درجة حريته (2) , وبفرض أنّ U,V متغيران عشوائيان معرفان بالشكل:

$$U = (X - Y) \quad , \quad V = Y$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة لهذين المتغيرين العشوائيين U,V .

2 – أوجد دالة الكثافة الاحتمالية الهامشية للمتغير العشوائي U .

الحل:

1 – بما أن X,Y متغيرين عشوائيين مستقلين, فإن :

$$f(x,y) = f(x)f(y)$$

$$= \left[\frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}x^{\frac{m}{2}-1}e^{-\frac{1}{2}x}\right]\left[\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}\right]$$

$$= \left[\frac{1}{2}e^{-\frac{1}{2}x}\right]\left[\frac{1}{2}e^{-\frac{1}{2}y}\right]$$

$$= \frac{1}{4}e^{-\left(\frac{x+y}{2}\right)} \;;\quad x,y > 0$$

بفرض أنّ:

$$\left.\begin{array}{c} u = (x-y) \\ v = y \end{array}\right\} \Longrightarrow \left\{\begin{array}{c} x = u+v \\ y = v \end{array}\right.$$

وبفرض أن هذا التحويل هو تحويل مطابق من $\mathcal{A} = \{(x,y); 0 < x < \infty , 0 < y < \infty \}$ إلى

$$\mathcal{B} = \{(u,v)\,; -u < v \;, 0 < v, -\infty < u < \infty\}$$

عندئذ فإن محدد اليعقوبي يعطى بالشكل التالي:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \Longrightarrow |J| = 1$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة لـ U,V تعطى بالشكل:

$$g(u,v) = f(x,y)|J| = \frac{1}{4}e^{-\left(\frac{x+y}{2}\right)}|1|\Bigg|_{\substack{x=u+v \\ y=v}}$$

$$= \frac{1}{4}e^{-\frac{1}{2}(u+2v)}$$

أي أنّ :

$$g(u,v) = \frac{1}{4}e^{-\frac{1}{2}(u+2v)} \quad ; \quad (u,v) \in \mathcal{B}$$

2 – لنوجد دالة الكثافة الهامشية للمتغير العشوائي U :

$$g(u) = \begin{cases} \int_{-u}^{\infty}\left(\frac{1}{4}e^{-\frac{1}{2}(u+2v)}\right)dv = \frac{1}{4}e^{\frac{1}{2}u} \;\; ; \; -\infty < u < 0 \\ \int_{0}^{\infty}\left(\frac{1}{4}e^{-\frac{1}{2}(u+2v)}\right)dv = \frac{1}{4}e^{-\frac{1}{2}u} \;\; ; \;\; 0 \le u < \infty \end{cases}$$

أي أنّ :

$$g(u) = \frac{1}{4}e^{-\frac{1}{2}|u|} \quad ; \; -\infty < \; u < \infty$$

مثال: لتكن X_1, X_2 عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي له التوزيع المنتظم المعرف على الفترة $[0,1]$, دالة كثافته الاحتمالية معرفة بالشكل:

$$f_X(x) = 1 \qquad ; 0 < x < 1$$

وبفرض أنّ Y_1 , Y_2 متغيرين عشوائيين معرفين بالشكل:

$$Y_1 = X_1 + X_2 \quad , \quad Y_2 = X_1 - X_2$$

المطلوب:

1 – أوجد دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين Y_1 , Y_2.

2 – أوجد دالة الكثافة الاحتمالية الهامشية لكل من المتغيرين العشوائيين Y_1 , Y_2 .

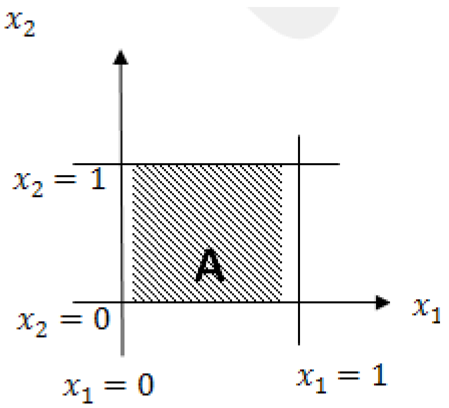
الحل:

1 – بما أن X_1, X_2 مستقلين فإن:

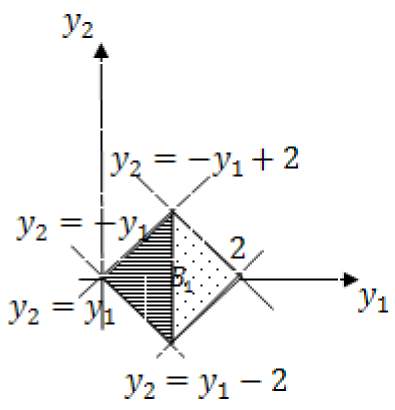
$$f_{X_1,X_2}(x_1,x_2) \overset{\text{من الاستقلال}}{\triangleq} f_{X_1}(x_1) \,.\, f_{X_2}(x_2) = 1 \;\; ; \;\; 0 < x_1,x_2 < 1$$

ولدينا من الفرض:

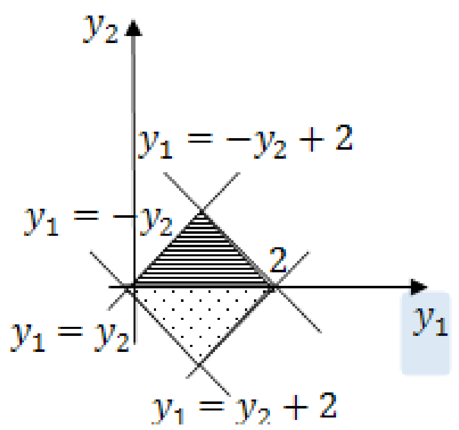
$$\left.\begin{array}{l} y_1 = x_1 + x_2 \\ y_2 = x_1 - x_2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} x_1 = \frac{y_1+y_2}{2} \\ x_2 = \frac{y_1-y_2}{2} \end{array}\right.$$



الشكل (1)



الشكل (2)



الشكل (3)

لنوجد محدد اليعقوبي:

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

وبالتالي فإن دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين Y_1 , Y_2 تعطى بالعلاقة التالية:

$$g_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2)\ \left|-\frac{1}{2}\right|\Bigg|_{\substack{x_1=\frac{y_1+y_2}{2} \\ x_2=\frac{y_1-y_2}{2}}} = 1 \times \frac{1}{2} = \frac{1}{2}$$

أي أن دالة الكثافة تصبح بالشكل:

$$g_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2} \quad ; (y_1,y_2)\epsilon\ B$$

2 – لنوجد دالة الكثافة الهامشية للمتغير العشوائي Y_1 :

$$g_{Y_1}(y_1) = \int_{R(y_2)} g_{Y_1,Y_2}(y_1,y_2)\ dy_2$$
$$= \left\{\begin{matrix} \int_{y_2=-y_1}^{y_2=y_1} \frac{1}{2} dy_2 \ = y_1 \,; 0 < y_1 \le 1 \\ \int_{y_2=y_1-2}^{y_2=-y_1+2} \frac{1}{2} dy_2 \ = 2 - y_1 \,; 1 < y_1 < 2 \end{matrix}\right\} (\text{شكل } 1)$$

نلاحظ أنّ :

$$\int_{-\infty}^{\infty} g_{Y_1}(y_1)dy_1 = \int_0^1 y_1 dy_1 + \int_1^2 (2-y_1)dy_1$$
$$= \left[\frac{y_1^2}{2}\right]_0^1 + \left[2y_1 - \frac{y_1^2}{2}\right]_1^2$$
$$= \left[\left(\frac{1}{2}\right) - (0)\right] + \left[(4-2) - \left(2-\frac{1}{2}\right)\right] \equiv 1$$

لنوجد دالة الكثافة الهامشية للمتغير العشوائي Y_2

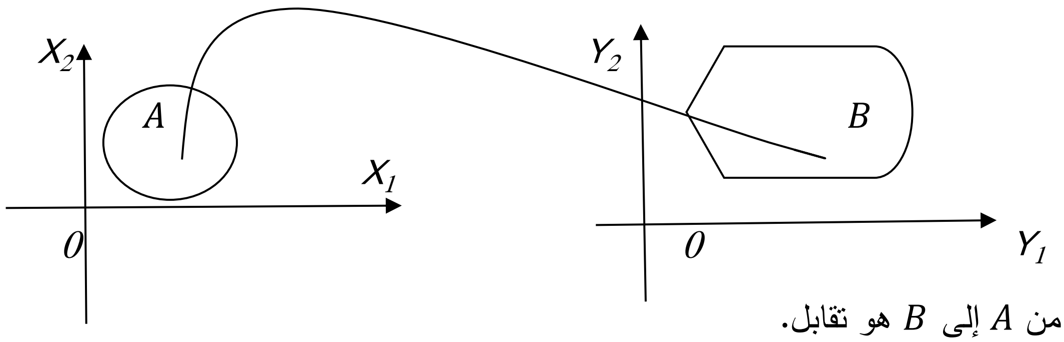
$$g_{Y_2}(y_2) = \int_{R(y_1)} g_{Y_1,Y_2}(y_1,y_2)\ dy_1$$
$$= \left\{\begin{matrix} \int_{y_1=-y_2}^{y_1=y_2+2} \frac{1}{2} dy_1 \ = 1 + y_2 \,; -1 < y_2 \le 0 \\ \int_{y_1=y_2}^{y_1=-y_2+2} \frac{1}{2} dy_1 \ = 1 - y_2 \,; 0 < y_2 < 1 \end{matrix}\right\} (\text{شكل } 2)$$

نلاحظ أنّ :

$$\int_{-\infty}^{\infty} g_{Y_2}(y_2)dy_2 = \int_{-1}^{0} (1+y_2)dy_2 + \int_0^1 (1-y_2)dy_2$$
$$= \left[y_2 + \frac{y_2^2}{2}\right]_{-1}^0 + \left[y_2 - \frac{y_2^2}{2}\right]_0^1$$
$$= \left[(0) - \left(-1+\frac{1}{2}\right)\right] + \left[\left(1-\frac{1}{2}\right) - (0)\right] \equiv 1$$

إيجاد الكثافة المشتركة لمتغيرين عشوائيين تابعين لمتغيرين آخرين(التوزيعات المنقطعة):

إذا كانت الدالة$f(x_1,x_2)$ الكثافة المشتركة للمتغيرين العشوائيين المنقطعين X_1,X_2 حيث A هي مجموعة النقاط والتي هي عبارة عن أزواج مرتبة من الأعداد الحقيقية التي تجعل دالة الكثافة المشتركة غير سالبة أي أنّ $f(x_1,x_2) \geq 0$ ولتكن كل من $Y_1 = U_1(X_1,X_2)$, $Y_2 = U_2(X_1,X_2)$ دالتين تعرفإنّ التحويل واحد لواحد أي $A \longrightarrow B$ كما في الشكل التالي:



أي أنّ التطبيق من A إلى B هو تقابل.

$$\left.\begin{aligned} y_1 = u_1(x_1,x_2) \\ y_2 = u_2(x_1,x_2) \end{aligned}\right\} \Longrightarrow \begin{cases} x_1 = u_1^{-1}(y_1,y_2) \\ x_2 = u_2^{-1}(y_1,y_2) \end{cases}$$

أنّ الكثافة المشتركة للمتغيرين العشوائيين الجديدين : $Y_1 = U_1(X_1,X_2)$, $Y_2 = U_2(X_1,X_2)$

تعطى بالعلاقة التالية:

$$g(y_1,y_2) = f[u_1^{-1}(y_1,y_2), u_2^{-1}(y_1,y_2)] \qquad ; \qquad (y_1,y_2) \in B$$

حيث: $(x_1,x_2) = (u_1^{-1}(y_1,y_2), u_2^{-1}(y_1,y_2)$ هي نقطة وحيدة(أحادية).

مثال: ليكن X,Y متغيرين عشوائيين بواسونيين مستقلين وسيطاهما على الترتيب λ_1,λ_2 وبفرض أنّ U,V متغيرين عشوائيين معرفين بالشكل:

$$U = X + Y \quad , \quad V = Y$$

والمطلوب:

1- أوجد دالة الكثافة المشتركة لـ U,V .

2- أوجد دالة الكثافة للمتغير U واستنتج نوعه.

الحل:

بما أنّ X,Y مستقلان فإنّ: $f(x,y) = f(x)f(y) = \frac{\lambda_1^x}{x!} e^{-\lambda_1} \frac{\lambda_2^y}{y!} e^{-\lambda_2}$

أي أنّ : $f(x,y) = \frac{\lambda_1^x . \lambda_2^y}{x! y!} e^{-(\lambda_1+\lambda_2)}$

ومنه نستطيع أنّ نجري التحويل التالي: $v = y \quad ; \quad u = x + y\} \Rightarrow x = u - v \quad , y = v$

وبالتالي فإنّ دالة الكثافة المشتركة لـ U,V تعطى بالشكل التالي:

$$g(u,v)=\frac{\lambda_1^{u-v}\lambda_2^{v}}{(u-v)!.v!}e^{-(\lambda_1+\lambda_2)}\quad ,(u,v)\in B$$

علماّ بأنّ: $B=\{(u,v): v=0,1,2,\dots,u \quad ; \quad u=0,1,2,\dots\}$

لنوجد دالة الكثافة للمتغير U كما يلي:

$$g(u)=\sum_{v=0}^{u} g(u,v)=\frac{e^{-(\lambda_1+\lambda_2)}}{\mathrm{u}!}\sum_{v=0}^{u}\frac{u!}{(u-v)!.v!}\lambda_1^{u-v}\lambda_2^{v}$$

وبالتالي: $g(u)=\frac{(\lambda_1+\lambda_2)^{\mathrm{u}}}{\mathrm{u}!}e^{-(\lambda_1+\lambda_2)} \quad : \quad u=0,1,\dots.$

وهو توزيع بواسوني وسيطه$(\lambda_1+\lambda_2)$ وبذلك يتم المطلوب.

الفصل الثاني

التقدير النقطي

Point estimation

لنفرض أنّ إحصائياً يريد أنّ يختار إجراءً معيناً أو عملاً من بين الإجراءات أو الأعمال الممكنة ولنفرض أنّ العمل أو الإجراء المناسب يتوقف على وسيط مجهول θ, نسمي دالة الكثافة بدلالة الوسيط $f(x,\theta)$, يسمى الوسيط في بعض الأحيان مَعْلَمَة (معالم) يحدد في حالة معرفته دالة الكثافة $f(x,\theta)$ للمجتمع المدروس.

أنّ الطريقة المتبعة هي أنّ يلجأ الإحصائي إلى أخذ عينة عشوائية من المجتمع $f(x,\theta)$ وأنّ يتخذ قراره حول الإجراء على ضوء النتائج التي نحصل عليها من العيّنة ولمعرفة ذلك لا بد من تلخيص الخطوات المتبعة:

1- تعيين مجموعة تعريف قيم الوسيط المجهول θ ونسميها فضاء الوسيط ونرمز له بـ θ.

2- تعيين كل الإجراءات الممكنة التي يمكن اتخاذها ونسمي هذه المجموعة فضاء التقدير ونرمز له الرمز A .

3- اختيار دالة مثل d في قيم العينة ولتكن $a=d(x_1,x_2,\dots,x_n)$, حيث $a\in A$ تسمى هذه الدالة دالة التقدير.

مثال: ليكن $f(x,\mu)$ دالة كثافة للمتغير العشوائي الذي له التوزيع الطبيعي توقعه μ وتباينه σ^2 ولنفرض أنّنا أمام اختيار أحد عملين أو قرارين:

الأول: نرمز له بالرمز a_1 وهو $\mu<0$.

الثاني: نرمز له بالرمز a_2 وهو $\mu\geq 0$.

فضاء الوسيط هو $\theta=\{\mu:\mu\in R\}$, وفضاء التقدير هو $A=\{a:a=a_1,a_2\}$ إذاً فضاء التقدير يحتوي على عنصرين فقط. لنفرض أنّ دالة القرار هي d حيث نتبنى القرار a_1 إذا كان $\overline{x}<0$ (متوسط العينة) ونتبنى القرار a_2 إذا كان $\overline{x}\geq 0$, أي نأخذ عينة عشوائية $X_1,X_2,\dots,X_n$ حجمها n ثم نحسب متوسط العينة $\overline{x}$, ونقول أنّ $\mu<0$ إذا كان $\overline{x}<0$ و $\mu\geq 0$ إذا كان $\overline{x}\geq 0$, ولدينا في كل مسألة تقدير مجموعة من الدوال التي يمكن أنّ نستخدم كلاً منها كدالة قرار , والسؤال أيّها نختار؟.

إنّ من المناسب أنّ نفترض أنّ النتيجة المترتبة على اختيار دالة القرار d وبالتالي اتخاذ القرار a هو خسارة يمكن التعبير عنها على شكل دالة عددية غير سالبة في كلٍّ من الوسيط θ والقرار a ونرمز لها بالرمز $l(a,\theta)$, وقيمة هذه الدالة تتوقف على قيم العينة, فإذا كرّرنا التجربة فإنّنا نجد في كل مرّة قيمة لدالة الخسارة لذلك نعرف

دالة جديدة تسمى دالة المخاطرة ونرمز لها بالرمز $R(d,\theta)$ وتساوي إلى التوقع الرياضي لدالة الخسارة أي:

$$R(d,\theta)=E[l(a,\theta)]$$

ويمكن الآن اعتبار مسألة الاستقراء كأنها مسألة اختيار قاعدة للتقدير أو اختيار دالة القرار, بحيث تكون دالة المخاطرة R في نهايتها الصغرى.

تعريف دالة الخسارة: ليكن θ وسيطاً يأخذ قيمه في فضاء الوسيط Θ وليكن a العمل أو القرار النهائي الذي يأخذ قيمه في فضاء القرار A , فعندئذٍ نقول أنّ $l(a,\theta)$ هي دالة الخسارة إذا كانت دالة عددية في كل من a و θ وتحقق الشرطين التاليين:

1- $l(a,\theta)\geq 0$ من أجل جميع قيم $a\in A$ و $\theta\in\Theta$.

2- $l(a,\theta)=0$ من أجل أي قيمة لـ θ , بحيث توجد على الأقل , قيمة واحدة لـ $a\in A$.

إنَ قيمة a التي تجعل دالة الخسارة مساوية للصفر هي التي تمثل القرار الصحيح الموافق للوسيط θ . والشرطان يعنيان أن الدالة $l(a,\theta)$ دالة غير سالبة وأنها تساوي الصفر عندما نتخذ القرار أو الفعل الصحيح.

تعريف دالة المخاطرة: إنّ دالة المخاطرة الموافقة لدالة القرار d ودالة الخسارة $l(a,\theta)$ والتي نرمز لها بالرمز $R(d,\theta)$ هي بالتعريف التوقع الرياضي لدالة الخسارة وتكتب بالشكل:

$$R(d,\theta)=E[l(a,\theta)]$$
$$=\int_{R^n} l[d(x_1,x_2,\dots,x_n);\theta]f(x_1,x_2,\dots,x_n;\theta)dx_1dx_2\dots dx_n$$

علماً أن القرار $a=d(x_1,x_2,\dots,x_n)$ هو بالضبط قيمة دالة القرار d من أجل عينة عشوائية $X_1,X_2,\dots,X_n$ حجمها n مأخوذة من مجتمع إحصائي دالة كثافته $f(x,\theta)$.

مثال: لتكن $X_1\,,X_2$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N(\mu\,,1)$ وليكن T مقدِّر لـ μ , حيث $T=\frac{1}{4}X_1+\frac{3}{4}X_2$. والمطلوب إيجاد دالة المجازفة $R(d,\mu)$ وذلك بفرض أنّ دالة الخسارة معرفة بالشكل : $l(T,\mu)=3\mu^2(T-\mu)^2$.

الحل:

لنوجد الآن التوقع الرياضي والتباين للمقدِّر T بالشكل:

$$E(T)=E\left(\frac{1}{4}X_1+\frac{3}{4}X_2\right)=E\left(\frac{1}{4}X_1\right)+E\left(\frac{3}{4}X_2\right)$$
$$=\frac{1}{4}E(X_1)+\frac{3}{4}E(X_2)=\frac{1}{4}\mu+\frac{3}{4}\mu=\mu$$

أي أنّ المقدِّر T هو مقدِّر غير منحاز بالنسبة للوسيط μ .

لنحسب تباين المقدِّر T بالشكل:

$$Var(T) = Var\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = Var\left(\frac{1}{4}X_1\right) + Var\left(\frac{3}{4}X_2\right)$$

وذلك لأنّ المتغيرين العشوائيين مستقلان , وبالاستفادة من خواص التباين نجد:

$$Var(T) = \frac{1}{16}Var(X_1) + \frac{9}{16}Var(X_2) = \frac{1}{16}(1) + \frac{9}{16}(1) = \frac{10}{16}$$

ولدينا حسب تعريف دالة المجازفة أنّ:

$$R(d,\theta) = E[l(a,\theta)] = E[3\mu^2(T-\mu)^2] = 3\mu^2 E(T-\mu)^2$$
$$= 3\mu^2 Var(T) = 3\mu^2\left(\frac{10}{16}\right) = \frac{30}{16}\mu^2 = \frac{15}{8}\mu^2$$

التقدير النقطي: إذا كان القرار المطلوب هو إيجاد تقدير للوسيط المجهول θ والذي نرمز له بالرمز $\hat{\theta}$ يحل محل الوسيط المجهول θ فإنّ $l(\hat{\theta},\theta) = 0 \Leftrightarrow \hat{\theta} = \theta$. فإذا فرضنا بشكلٍ عام أنّ دالة الخسارة $l(\hat{\theta},\theta) = 0$ إذا وفقط إذا كانت $\hat{\theta} = \theta$, عندئذٍ نقول أنّ دالة الخسارة تحدد لنا ما يسمى بمسألة التقدير النقطي وتدعى دالة القرار $\hat{\theta} = d(x_1,x_2,\dots,x_n)$ مقدِّر θ , ونختار هذا القرار بحيث تكون المخاطرة أقل ما يمكن.

مفهوم الإحصاء الكافي: أنّ الطريقة الأساسية لأخذ معلومات عن مجتمع إحصاء كافي هو أنّ نأخذ عينة عشوائية منه ونقوم بدراسة هذه العينة, وعلى أساس هذه الدراسة نستخلص كل المعلومات المطلوبة عن هذا المجتمع وقد وجد بعض الإحصائيين أنّ هناك بعض الدوال تشمل علة جميع الصفات التي يجب أنّ نعرفها من هذه العينة عن المجتمع المدروس, فمثلاً الدالة: $Y = X_1$ لا تعطينا كل المعلومات عن المجتمع المأخوذ منه تلك العينة, عكس ذلك الأمر بالنسبة للدالة $Y = minX_i$, بينما تعطينا الدالة $Y = X_1 + X_2 + \cdots + X_n$ كل المعلومات حول المجتمع الذي أخذنا منه العينة العشوائية, لذلك يكون هدفنا هو دراسة مثل هذه الإحصاءات ومن ثم تعميمها على المجتمع.

لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي قانونه $f(x,\theta)$ عندئذٍ تكون كثافة هذه العينة $L(X_1,X_2,\dots,X_n,\vartheta) = \prod_{i=1}^{n} f(x_i,\theta)$ فإذا أخذنا الإحصاءات التالية:

$$\left.\begin{array}{l} Y_1 = U_1(X_1,X_2,\dots,X_n) \\ Y_2 = U_2(X_1,X_2,\dots,X_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \\ Y_n = U_n(X_1,X_2,\dots,X_n) \end{array}\right\} \qquad (1)$$

فإنّ التحويلات المناسبة, (قيم هذه الدوال):

$$\left.\begin{array}{l} y_1 = u_1(x_1,x_2,\dots,x_n) \\ y_2 = u_2(x_1,x_2,\dots,x_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \\ y_n = u_n(x_1,x_2,\dots,x_n) \end{array}\right\} \qquad (1)$$

لنفرض أنّ هذه التحويلات أحادية, (أي يوجد تقابل) عندئذٍ تكافئ جملة المعادلات (1) جملة المعادلات التالية:

$$\left.\begin{array}{l} x_1 = u_1^{-1}(y_1, y_2, \dots, y_n) \\ x_2 = u_2^{-1}(y_1, y_2, \dots, y_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_n = u_n^{-1}(y_1, y_2, \dots, y_n) \end{array}\right\} \qquad (2)$$

فإذا كانت هذه الحلول وحيدة فعندئذٍ تكون الكثافة المشتركة لـ: $Y_1, Y_2, \dots, Y_n$ معطاة بالشكل:

$$g(y_1, y_2, \dots, y_n, \theta) = f\left[u_1^{-1}\left(\underline{y}, \theta\right), u_2^{-1}\left(\underline{y}, \theta\right), \dots, u_n^{-1}\left(\underline{y}, \theta\right), \theta\right] |J|$$

حيث J هو معين اليعقوبي:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots\dots\dots & \frac{\partial x_1}{\partial u_r} & \frac{\partial x_1}{\partial u_{r+1}} & \dots\dots\dots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots\dots\dots & \frac{\partial x_2}{\partial u_r} & \frac{\partial x_2}{\partial u_{r+1}} & \dots\dots\dots & \frac{\partial x_2}{\partial u_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \dots\dots\dots & \frac{\partial x_n}{\partial u_r} & \frac{\partial x_n}{\partial u_{r+1}} & \dots\dots\dots & \frac{\partial x_n}{\partial u_n} \end{vmatrix}$$

وبالتالي نستطيع إيجاد الكثافة الشرطية $h(y_2, y_3, \dots, y_n \backslash y_1, \theta)$ بالعلاقة التالية:

$$h(y_2, y_3, \dots, y_n \backslash y_1, \theta) = \frac{g(y_1, y_2, \dots, y_n, \theta)}{g_1(y_1, \theta)}$$

حيث $g_1(y_1, \theta)$ هي الكثافة الهامشية للمتغير Y_1 وتكون فقط تابعة لـ Y_1 والوسيط θ, فإذا كانت الكثافة الشرطية لا تتعلق بالوسيط θ فإنّ الكثافة الشرطية تمثل إحصاء وعندئذٍ لا يمكن أنّ نعرف أي شيء عن الوسيط المجهول θ.

تعريف الإحصاء الكافي: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه $f(x, \theta)$ عندئذٍ نقول عن الإحصاء $Y_1 = U_1(X_1, X_2, \dots, X_n)$ أنّه إحصاء كافي بالنسبة لـ θ إذا كانت الكثافة الشرطية: $h(y_2, y_3, \dots, y_n \backslash y_1, \theta)$ لا تتعلق بالوسيط θ (مستقلة عن θ) وذلك أياً كانت الإحصاءات: $Y_i = U_i(X_1, X_2, \dots, X_n)$ حيث $i = 2,3, \dots, n$:

مبرهنة: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه $f(x, \theta)$ فإنّ الشرط اللازم والكافي لكي يكون الإحصاء $Y_1 = U_1(X_1, X_2, \dots, X_n)$ إحصاء كافي بالنسبة لـ θ هو أنّ يتحقق الشرط التالي:

$$f(x_1, x_2, \dots, x_n, \theta) = \underbrace{g_1(y_1, \theta)}_{\text{دالة الكثافة الهامشية لـ} Y_1} \; \underbrace{H(x_1, x_2, \dots, x_n)}_{\text{لا يتعلق بـ } \theta} \qquad (1)$$

البرهان:

1 – لزوم الشرط : لنفرض أنّ الشرط (1) محقق ولنثبت أنّ Y_1 هو إحصاء كافي بالنسبة لـ θ .

لتكن $i=2,3,\dots,n$, $Y_i=U_i(X_1,X_2,\dots,X_n)$, إحصاءات اختيارية وبالتالي فإنّ الكثافة المشتركة لهذه المتغيرات العشوائية تعطى بالشكل: $g(y_1,y_2,\dots,y_n,\theta)=f(x_1,x_2,\dots,x_n,\theta)|J|$ وذلك بتبديل $x_1,x_2,\dots,x_n$ بقيمها من التحويلات المماثلة.

$$g(y_1,y_2,\dots,y_n,\theta)=f\left[u_1^{-1}\left(\underline{y},\theta\right),u_2^{-1}\left(\underline{y},\theta\right),\dots,u_n^{-1}\left(\underline{y},\theta\right),\theta\right]|J|$$

وبما أنّ الشرط (1) محقق فإنّ:

$$g(y_1,y_2,\dots,y_n,\theta)=g_1(y_1,\theta)H(x_1,x_2,\dots,x_n)|J|$$

ولكن الكثافة الشرطية h تعطى بالعلاقة:

$$\begin{aligned}h(y_2,y_3,\dots,y_n\backslash y_1,\theta)&=\frac{g(y_1,y_2,\dots,y_n,\theta)}{g_1(y_1,\theta)}\\&=\frac{g_1(y_1,\theta)H(x_1,x_2,\dots,x_n)|J|}{g_1(y_1,\theta)}=H(x_1,x_2,\dots,x_n)|J|\end{aligned}$$

وبما أنّ H مستقل عن θ وكذلك معين اليعقوبي, فإنّ الكثافة الشرطية لا تتعلق بـ θ وبالتالي يكون Y_1 إحصاءً كافياً لـ θ.

2 – كفاية الشرط: نفرض أنّ Y_1 إحصاء كافي بالنسبة لـ θ ولنثبت أنّ الشرط (1) محقق:

بما أنّ Y_1 إحصاءً كافياً بالنسبة لـ θ فإنّ:

$$g(y_1,y_2,\dots,y_n,\theta)=g_1(y_1,\theta)h(y_2,y_3,\dots,y_n\backslash y_1)$$

لكن:

$$f(x_1,x_2,\dots,x_n,\theta)=f(x_1,\theta).f(x_2,\theta)\dots f(x_n,\theta)=g(y_1,y_2,\dots,y_n,\theta)|J^*|$$

حيث: J^* هو التحويل المعاكس من X إلى Y وبالتالي:

$$f(x_1,x_2,\dots,x_n,\theta)=g_1(y_1,\theta)h(y_2,y_3,\dots,y_n\backslash y_1)|J^*|$$

حيث: J^*,h غير تابع لـ θ, أي أنّ:

$$f(x_1,x_2,\dots,x_n,\theta)=g_1(y_1,\theta)H(x_1,x_2,\dots,x_n)$$

أي أنّ الشرط (1) محقق.

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه:

$$f(x,p)=p^x(1-p)^{1-x}\qquad,\qquad x\geq 0\ ,0<p<1$$

أثبت أنّ الإحصاء $Y_1=X_1+X_2+\dots+X_n$ هو إحصاء كافي بالنسبة لـ p.

<u>الحل</u>: من أجل ذلك لنثبت أنّ الشرط (1) محقق أي:

$$f(x_1,x_2,\dots,x_n,\theta)=\underbrace{g_1(y_1,\theta)}_{Y_1\text{ دالة الكثافة الهامشية لـ}}\ \underbrace{H(x_1,x_2,\dots,x_n)}_{\theta\text{ لا يتعلق بـ }}$$

لنوجد دالة الكثافة المشتركة للمتغيرات العشوائية $X_1,X_2,\dots,X_n$ بالشكل:

$$f(x_1,x_2,\dots,x_n,\theta)=f(x_1,\theta).f(x_2,\theta)\dots f(x_n,\theta)$$

$$=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$$

$$=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}$$

$$=\binom{n}{\sum_{i=1}^{n}x_i}p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}\frac{1}{\binom{n}{\sum_{i=1}^{n}x_i}}$$

$$=\binom{n}{\sum_{i=1}^{n}x_i}p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}\frac{(\sum_{i=1}^{n}x_i)!(n-\sum_{i=1}^{n}x_i)!}{n!}$$

$$=\binom{n}{y_1}p^{y_1}(1-p)^{n-y_1}\frac{(\sum_{i=1}^{n}x_i)!(n-\sum_{i=1}^{n}x_i)!}{n!}$$

$$=\underbrace{g_1(y_1,\theta)}_{Y_1\text{ دالة الكثافة الهامشية لـ}}\quad\underbrace{H(x_1,x_2,\dots,x_n)}_{\theta\text{ لا يتعلق بـ}}$$

حيث $Y_1=\sum_{i=1}^{n}X_i$, وبالتالي Y_1 هو إحصاءً كافياً , لأنّ المقدار $\binom{n}{y_1}p^{y_1}(1-p)^{n-y_1}$ هو الكثافة الهامشية لـ Y_1 والمقدار الثاني $\frac{(\sum_{i=1}^{n}x_i)!(n-\sum_{i=1}^{n}x_i)!}{n!}$ لا يتعلق بـ θ .

ملاحظة: قد يكون الشكل التحليلي لدالة ما مستقلة عن متغير ما, لكن لكي يكون مستقل فعلاً عن ذلك المتغير, يجب أنّ يكون مستقل بشكله التحليلي بالإضافة إلى تعيين مجموعة التعريف.

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه:

$$f(x,\theta)=e^{-(x-\theta)}\quad,\quad\theta<x<\infty\quad,\quad-\infty<\theta<\infty$$

أثبت أنّ الإحصاء $Y_1=\min_i X_i$ هو إحصاء كاف بالنسبة لـ θ .

الحل:

لنوجد الكثافة الهامشية لـ Y_1.

$$g_1(y_1,\theta)=\frac{d}{dy_1}G_1(y_1,\theta)$$

لكن :

$$G_1(y_1,\theta)=P(Y_1<y_1)=1-P(Y_1>y_1)$$

$$\Rightarrow P[Y_1>y_1]=P[X_1>y_1]\dots..P[X_n>y_1]=\{P[X_1>y_1]\}^n$$

لأنّ العينة عشوائية فالمتغيرات مستقلة.

$$P[X_1>y_1]=\int_{y_1}^{\infty}e^{-(x-\theta)}dx=e^{-\theta}.(-e)^{-x}=e^{\theta}e^{-y_1}=e^{-(y_1-\theta)}$$

$$\Rightarrow P(Y_1>y_1)=\left[e^{-(y_1-\theta)}\right]^n=e^{-n(y_1-\theta)}$$

عندئذٍ :

$$G_1(y_1,\theta)=1-e^{-n(y_1-\theta)}$$

وبالاشتقاق بالنسبة لـ y_1 نجد:

$$g_1(y_1,\theta)=\frac{dG_1(y_1,\theta)}{dy_1}=ne^{-n(y_1-\theta)}=ne^{-n\,min\,x_i}.e^{n\theta}$$

لنرى فيما إذا كانت النظرية محققة أم لا:

$$f(x_1,x_2,\dots,x_n,\theta)=e^{-(x_1-\theta)}e^{-(x_2-\theta)}\dots e^{-(x_n-\theta)}$$
$$=e^{-\sum(x_i-\theta)}=e^{-\sum x_i}.e^{n\theta}$$
$$=e^{-\sum x_i}e^{n\theta}.\frac{ne^{-n\,min\,x_i}}{ne^{-n\,min\,x_i}}=ne^{-n(min\,x_i-\theta)}.\frac{e^{-\sum x_i}}{ne^{-n\,min\,x_i}}$$
$$=g_1(y_1,\theta).H(x_1,x_2,\dots,x_n)$$

مبرهنة: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه $f(x,\theta)$ وليكن $Y_1=U_1(X_1,X_2,\dots,X_n)$ إحصاءً ما, الشرط اللازم والكافي كي يكون الإحصاء Y_1 إحصاءً كافياً بالنسبة لـ θ هو تحقق الشرط :

$$f(x_1,x_2,\dots,x_n,\theta)=K_1(y_1,\theta).K_2\left(\underline{x_1,x_2,\dots,x_n}\right)\dots(**)$$

حيث K_1 تابع لـ $x_1,x_2,\dots,x_n$ عن طريق y_1 وتابع للوسيط θ بينما K_2 تابع فقط لـ $x_1,x_2,\dots,x_n$.

البرهان (لزوم الشرط) : نفرض أنّ الشرط $(**)$ محقق ولنثبت أنّ Y_1 إحصاء كافي.

لنفرض أنّ الإحصاء $Y_k=U_k(X_1,\dots,X_n)$ حيث $k=2,\dots,n$ ولنفرض أنّ:

$$g(y_2,y_3,\dots,y_n,\theta)=f(x_1,x_2,\dots,x_n,\theta)|J|=$$
$$K_1(y_1,\theta).K_2[w_1(y_2,y_3,\dots,y_n)],\dots,K_n[w_n(y_2,y_3,\dots,y_n)]|J|$$
$$=K_1(y_1,\theta).[m(y_2,y_3,\dots,y_n)]$$
$$g_1(y_1,\theta)=\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}K_1(y_1,\theta).[m(y_2,y_3,\dots,y_n)]\,dy_2\dots dy_n=K_1(y_1,\theta).\left(m(y_1)\right)$$

لنطبق التعريف الأساسي للإحصاء الكافي (أي الكافة الشرطية لا تتعلق بـ θ) أي:

$$h(y_2,\dots,y_n|y_1)=\frac{g(y_2,y_3,\dots,y_n,\theta)}{g_1(y_1,\theta)}=\frac{K_1(y_1,\theta).(m(y_2,y_3,\dots,y_n))}{K_1(y_1,\theta).(m(y_1))}=\frac{m(y_2,y_3,\dots,y_n)}{m(y_1)}$$

والمقدار في الطرف الأيمن لا يتعلق بـ θ إذاً الإحصاء هو إحصاء كاف.

(كفاية الشرط) : لنفرض أنّ Y_1 هو إحصاء كافٍ ولنثبت أنّ الشرط $(**)$ محقق بتطبيق المبرهنة السابقة.

$$f(x_1,x_2,\dots,x_n,\theta)=g_1(y_1,\theta).H(x_1,x_2,\dots,x_n)=K_1(y_1,\theta).K_2(x_1,x_2,\dots,x_n)$$

إذا الشرط $(**)$ محقق.

مثال: لو عدنا إلى المثال الأول, فإنّنا نجد:

$$f(x_1,x_2,\dots,x_n,\theta)=p^{\sum x_i}(1-p)^{n-\sum x_i}=K_1(y_1,\theta).K_2(x_1,x_2,\dots,x_n)$$

حيث $K_2=1$ و K_1 تابع لـ $x_1,x_2,\dots,x_n$ عن طريق التحويل $y_1=\sum x_i$ وتابع لـ θ.

مثال: لو عدنا إلى المثال السابق (الثاني) فقد وجدنا أنّ الكثافة المشتركة:

$$f(x_1, x_2, \dots, x_n, \theta) = e^{-\sum x_i + n\theta} . \frac{\min x_i}{\min x_i}$$
$$= e^{n\theta} (\min x_i) \frac{e^{-\sum x_i}}{\min x_i} = K_1(\min x_i , \theta) K_2(x_1, x_2, \dots , x_n)$$

إذاً Y_1 هو إحصاء كاف بالنسبة لـ θ.

مثال: لتكن $X_1, X_2, \dots , X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي دالة كثافته معرفة بالشكل :

$$f(x, \theta) = \theta x^{\theta - 1} \ , \quad (0 < x < 1)$$

أثبت أنّ $Y_1 = X_1 X_2 \dots X_n = \prod_{i=1}^{n} X_i$ إحصاء كاف بالنسبة لـ θ.

الحل:

$$f(\underline{x}, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta)$$
$$= \prod_{i=1}^{n} \left(\theta x_i^{\theta - 1}\right) = \theta^n (x_1\ x_2 \dots\ x_n)^{\theta - 1}$$
$$= \theta^n (x_1\ x_2 \dots\ x_n)^{\theta}\ \frac{1}{x_1\ x_2 \dots x_n}$$
$$= K_1(u_1(x_1, x_2, \dots , x_n), \theta) . K_2(x_1, x_2, \dots , x_n)$$

تمرين: لتكن $X_1, X_2, \dots , X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه $\mu = 0$ و $\sigma^2 = \theta$ أ ي أنّ $f(x, \theta) \sim N(0, \sigma^2 = \theta)$ والمطلوب :

أثبت أنّ $Y_1 = X_1^2 + X_2^2 + \dots + X_n^2$ إحصاء كاف بالنسبة لـ θ.

التقديرات غير المنحازة:

إذا كان المقدار $(\hat{\theta} - \theta)$ هو انحراف $\hat{\theta}$ عن الوسيط. فتسمى الدالة $(\hat{\theta} - \theta)^2$ بالخطأ التربيعي ويسمى التوقع الرياضي لـ $(\hat{\theta} - \theta)^2$ أي $E\left[(\hat{\theta} - \theta)^2\right]$ توقع الخطأ التربيعي, أو متوسط الخطأ التربيعي . ويكتب متوسط الخطأ التربيعي بشكل آخر:

$$E\left[\left(\hat{\theta} - \theta\right)^2\right] = E\{[\hat{\theta} - E(\hat{\theta})] - [\theta - E(\hat{\theta})]\}^2$$
$$= E\{[\hat{\theta} - E(\hat{\theta})]\}^2 + E\{[\theta - E(\hat{\theta})]\}^2$$
$$= Var\ \hat{\theta} + E\{[\theta - E(\hat{\theta})]\}^2$$

أي أنّ متوسط الخطأ التربيعي هو عبارة عن مجموع كميتين غير سالبتين أولهما تباين التقدير $\hat{\theta}$ والثاني هو انحراف التقدير ويمكن أنّ يكون الانحراف موجباً أو سالباً أو معدوماً وإذا استطعنا الحصول على تقدير انحرافه قريب من الصفر وتباينه صغير جداً فسيكون متوسط الخطأ التربيعي صغير ومن الواضح أنّنا نفضل أنّ يكون الانحراف معدوماً وفي هذه الحالة يكون التقدير غير منحازاً.

تعريف التقدير غير المنحاز:

نقول عن التقدير $\hat{\theta}$ أنّه غير منحاز بالنسبة إلى θ إذا كان توقعه الرياضي مساوياً الوسيط نفسه , أي $E(\hat{\theta}) = \theta$ وذلك من أجل جميع قيم θ المنتهية إلى فضاء الوسيط θ .

مثال: لتكن $X_1, \ldots, X_n$ عينة عشوائية حجمها n متوسطها μ مأخوذة من مجتمع إحصائي قانونه $f(x, \mu)$ عندئذٍ يكون: $E(\overline{X}) = \mu$ أي أنّ $(\overline{X})$ هو تقدير غير منحاز بالنسبة لـ μ.

$$E(\overline{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}(EX_1 + \cdots + EX_n) = \frac{n\mu}{n} = \mu$$

أي أنّ $\overline{X}$ هو تقدير غير منحاز بالنسبة لـ μ ويكون أيضاً X_1 هو تقدير غير منحاز بالنسبة لـ μ لأنّ $EX_1 = \mu$.

التقدير المتّسق: يبدو لنا من تعريف العينة العشوائية أنّ التقدير المبني على أساس عينة حجمها n ملاحظة,(أو n قياس), يجب أنّ يكون أفضل من التقدير المبني على أساس عينة تحوي أقل من n قياساً.

لتوضيح ذلك نفرض أنّ θ_1 هو تقدير للوسيط θ مأخوذ من عينة عشوائية حجمها $n = 1$ ملاحظة. و θ_2 هو تقدير للوسيط θ مأخوذ من عينة عشوائية حجمها $n = 2$ ملاحظة, وهكذا... أي لتكن $\{\theta_n\}$ متتالية من التقديرات للوسيط θ وأنّ هذه المتتالية من التقديرات تعني أنّ متتالية التقديرات $\{\theta_n\}$ تتقارب احتمالياً من θ أو تنتهي احتمالياً إلى θ عندما n تكون كبيرة: $n \to \infty$ ومنه نستنتج التعريف التالي:

تعريف: ليكن $\hat{\theta}_n$ مقدّراً مبنياً على عينة حجمها n . نقول أنّ المقدّر $\hat{\theta}_n$ أنّه مقدِّر متّسق للوسيط θ إذا وفقط إذا كان لدينا من أجل أي $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = 0$$

أو بشكل مكافئ : $\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \le \varepsilon\right) = 1$

أي أنّه من أجل أي عدد موجب ε مهما كان صغيراً يمكن التحكم باحتمال ألا يتجاوز الخطأ المطلق للتقدير العدد ε , وجعل هذا الاحتمال قريباً من الواحد بقدر ما نريد وذلك بزيادة حجم العينة n .

وفي الحالة الخاصة التي يكون فيها $\hat{\theta}_n = \bar{X}$ متوسط العينة كمقدِّر لـ $\theta = \mu$ متوسط المجتمع الذي جاءت منه العينة , يصبح هذا التعريف مطابقاً لقانون الأعداد الكبيرة .

مبرهنة: إذا كان $E\left(\hat{\theta}_n - \theta\right)^2 \xrightarrow[n\to\infty]{} 0$ فإنّ $\hat{\theta}_n$ يكون تقديراً متّسقاً لـ θ.

البرهان: لدينا بفرض $\varepsilon > 0$ عدد موجب.

$$E(\hat{\theta}_n - \theta)^2 = \int_{-\infty}^{+\infty} \left(\hat{\theta}_n - \theta\right)^2 f(\theta)d\theta$$

$$= \int_{-\infty}^{\theta-\varepsilon} \left(\hat{\theta}_n - \theta\right)^2 f(\theta)d\theta + \int_{\theta-\varepsilon}^{\theta+\varepsilon} \left(\hat{\theta}_n - \theta\right)^2 f(\theta)d\theta + \int_{\theta+\varepsilon}^{+\infty} \left(\hat{\theta}_n - \theta\right)^2 f(\theta)d\theta$$

لدينا في التكاملين الأول والثالث $\left(\hat{\theta}_n-\theta\right)^2>\varepsilon^2$ ولدينا التكامل الثاني غير سالب, عندئذٍ تصبح العلاقة بالشكل:

$$E\left(\hat{\theta}_n-\theta\right)^2 \geq \int_{-\infty}^{\theta-\varepsilon} \varepsilon^2 f(\theta)d\theta+\int_{\theta+\varepsilon}^{+\infty} \varepsilon^2 f(\theta)d\theta$$
$$=\ \varepsilon^2\left\{p\left(\hat{\theta}_n \leq \theta-\varepsilon\right)+p\left(\hat{\theta}_n>\theta+\varepsilon\right)\right\}=\varepsilon^2 p\left(\left|\hat{\theta}_n-\theta\right|>\varepsilon\right)$$

نقسم الطرفين على $\varepsilon^2>0$ فنجد:

$$p\left(\left|\hat{\theta}_n-\theta\right|>\varepsilon\right) \leq \frac{E\left(\hat{\theta}_n-\theta\right)^2}{\varepsilon^2}$$

حسب شروط المبرهنة بما أنّ الطرف الأيمن يسعى إلى الصفر فإنّ :

$$p\left(\left|\hat{\theta}_n-\theta\right|>\varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

وهذا يعني أنّ $\hat{\theta}_n$ هو تقدير متّسق وذلك حسب التعريف السابق.

نتيجة: إذا كان θ_n تقديراً غير منحازاً لـ θ فإنّ نهاية المقدار $E\left(\hat{\theta}_n-\theta\right)^2$ تسعى إلى الصفر عندما $n\to\infty$ تعني أنّ تباين $\hat{\theta}_n$ يسعى إلى الصفر عندما $n\to\infty$. إذاً فالتقدير غير المنحاز والذي يسعى تباينه إلى الصفر عندما تسعى إلى اللانهاية , يكون تقديراً متّسقاً أيضاً. وبعبارة أخرى :

ε عدد موجب محدود , $p\left(\left|\hat{\theta}_n-\theta\right|>\varepsilon\right) \leq \frac{Var\ (\hat{\theta}_n)}{\varepsilon^2}$

التقدير غير المنحاز ذو التباين الأصغري: أنّ مقارنة تقديرين من خلال مقارنة متوسط الخطأ التربيعي لكل أنّهما هي طريقة مفضلة للحكم على التقدير الأجود من بين مجموعة التقديرات المتوفرة لنا, إلا أنّه نادراً ما نقع على تقدير يتصف بأنّ متوسط خطأه التربيعي أصغر ما يمكن. فعندئذٍ نلجأ إلى أساليب أخرى.

تعريف: نقول عن $\hat{\theta}$ أنّه تقدير غير منحاز ذو تباين أصغري للوسيط θ إذا تحقق:

$E\left(\hat{\theta}\right)=\theta$ و $Var\left(\hat{\theta}\right)$ هو أصغر من تباين أي تقدير غير منحاز آخر لـ θ.

ملاحظة: إذا كان تساوق التقدير صفة هامة من أجل العينات كبيرة الحجم فإنّ الصفة التي عرفناها في الفقرة السابقة لها أهميتها بصرف النظر عن حجم العينة ولها أهميتها الخاصة بالنسبة للعينات ذات الأحجام الصغيرة.

وللوصول إلى تقدير غير منحاز ذو تباين أصغري نبدأ بإيجاد حد أدنى للتباين ضمن صف التقديرات غير المنحازة للوسيط θ فإذا كان تباين التقدير غير المنحاز $\hat{\theta}$ مثلاً محققاً بشرط الحد الأدنى مهما كان θ قلنا عنه أنّه التقدير المطلوب غير المنحاز ذو التباين الأصغري.

فعالية تقدير: إذا كان T_1, T_2 مقدّرين غير منحازين للوسيط θ , فعندئذ نقول إن المقدِّر T_1 أفضل أو أجود أو أكثر فعالية من المقدِّر T_2 إذا كان تباين المقدِّر T_1 أصغر من تباين المقدِّر T_1 , أي إذا تحققت العلاقة:

$$Var(T_1) \leq Var(T_2)$$

وبتقسيم طرفي المتباينة على العدد الموجب $Var(T_2)$ نجد أن :

$$\frac{Var(T_1)}{Var(T_2)} \leq 1$$

إن هذه المتباينة تدل على أن الفعالية تتناسب عكساً مع التباين. أي أنه كلما كان التباين أصغر , كلما كان المقدِّر أكثر فعالية.

كما وتستخدم عادة , عندما يكون حجم العينة كبيراً , النهاية $\lim_{n\to\infty}\left[\frac{Var(T_1)}{Var(T_2)}\right]$, في تعريف فعالية مقدِّر T_2 غير منحاز بالنسبة لمقدِّر T_1 غير منحاز آخر بالنسبة لنفس الوسيط.

مثال: لتكن X_1 , X_2 عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N(\mu , 1)$ ولتكن T_1, T_2, T_3 ثلاثة مقدِّرات لـ μ , معرفة بالعلاقات التالية: $T_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2$, $T_2 = \frac{1}{3}X_1 + \frac{2}{3}X_2$, $T_3 = \frac{1}{4}X_1 + \frac{3}{4}X_2$. وبفرض أنّ دالة الخسارة معرفة بالشكل : $\mathrm{I}(\mathrm{T}, \mu) = 3\mu^2(\mathrm{T} - \mu)^2$. والمطلوب:

1 – إيجاد دالة المجازفة $R(T, \mu)$ للمقدِّرات الثلاثة.

2 – بين أياً من المقدّرات أفضل.

الحل:

لنوجد الآن التوقع الرياضي والتباين للمقدِّرات الثلاثة T_1, T_2, T_3 بالشكل:

$$E(T_1) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = E\left(\frac{1}{2}X_1\right) + E\left(\frac{1}{2}X_2\right)$$
$$= \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

$$E(T_2) = E\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = E\left(\frac{1}{3}X_1\right) + E\left(\frac{2}{3}X_2\right)$$
$$= \frac{1}{3}E(X_1) + \frac{2}{3}E(X_2) = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu$$

$$E(T_3) = E\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = E\left(\frac{1}{4}X_1\right) + E\left(\frac{3}{4}X_2\right)$$
$$= \frac{1}{4}E(X_1) + \frac{3}{4}E(X_2) = \frac{1}{4}\mu + \frac{3}{4}\mu = \mu$$

أي أنّ المقدِّرات الثلاثة T_1, T_2, T_3 هي مقدِّرات غير منحازة بالنسبة للوسيط μ .

لنحسب تباينات هذه المقدِّرات الثلاثة T_1, T_2, T_3 بالشكل:

$$Var(T_1) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = Var\left(\frac{1}{2}X_1\right) + Var\left(\frac{1}{2}X_2\right)$$

وذلك لأنّ المتغيرين العشوائيين مستقلان , وبالاستفادة من خواص التباين نجد:

$$Var(T_1) = \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) = \frac{1}{4}(1) + \frac{1}{4}(1) = \frac{1}{2} = \frac{72}{144}$$

$$Var(T_2) = Var\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = Var\left(\frac{1}{3}X_1\right) + Var\left(\frac{2}{3}X_2\right)$$

وذلك لأنّ المتغيرين العشوائيين مستقلان , وبالاستفادة من خواص التباين نجد:

$$Var(T_2) = \frac{1}{9}Var(X_1) + \frac{4}{9}Var(X_2) = \frac{1}{9}(1) + \frac{4}{9}(1) = \frac{5}{9} = \frac{80}{144}$$

$$Var(T_3) = Var\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = Var\left(\frac{1}{4}X_1\right) + Var\left(\frac{3}{4}X_2\right)$$

وذلك لأنّ المتغيرين العشوائيين مستقلان , وبالاستفادة من خواص التباين نجد:

$$Var(T_3) = \frac{1}{16}Var(X_1) + \frac{9}{16}Var(X_2) = \frac{1}{16}(1) + \frac{9}{16}(1) = \frac{10}{16} = \frac{90}{144}$$

نلاحظ أن:

$$Var(T_1) = \frac{72}{144} < Var(T_2) = \frac{80}{144} < Var(T_3) = \frac{90}{144}$$

أي أن المقدِّر T_1 أفضل من المقدِّر T_2 وكلاهما أفضل من المقدِّر T_3 .

مثال: لتكن لدينا $X_1, X_2, \dots \dots, X_8$ عينة عشوائية حجمها $n = 8$ مأخوذة من مجتمع إحصائي الموصوف بدالة الكثافة :

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad ; x \in \mathcal{R}$$

ولنأخذ:

$$T_1 = a\{(X_1 - X_2)^2 + (X_3 - X_4)^2 + (X_5 - X_6)^2 + (X_7 - X_8)^2\}$$

مقدِّر للوسيط σ^2.

$$T_2 = \frac{1}{7}\sum_{i=1}^{8}(x_i - \bar{x})^2$$

مقدِّر للوسيط σ^2 .

المطلوب:

1 – أوجد قيمة الثابت a حتى يكون T_1 غير منحاز بالنسبة لـ σ^2.

2 – بين أي المقدِّرين أفضل .

الحل:

1 – لتعيين الثابت a بحيث يكون $E(T_1) = \sigma^2$, نقوم بدراسة المتغير العشوائي $Y = (X_1 - X_2)^2$.

ولدراسة هذا المتغير العشوائي ندرس المتغير $X_1 - X_2$. أنّ المتغير $X_1 - X_2$ هو متغير عشوائي طبيعي توقعه الرياضي صفر وتباينه $2\sigma^2$ وذلك لأنّ :

$$\varphi_{X_1-X_2}(t) = Ee^{t(x_1-x_2)} = E(e^{tx_1}.e^{tx_2}) = E(e^{tx_1}).E(e^{-tx_2})$$
$$= e^{\mu t+\frac{\sigma^2}{2}t^2}.e^{-\mu t+\frac{\sigma^2}{2}t^2} = e^{\sigma^2 t^2}$$

وهذا يدل على أنّ المتغير $X_1 - X_2$ هو متغير عشوائي طبيعي وسيطاه صفر و $2\sigma^2$.

لنأخذ الآن المتغير $\frac{X_1-X_2}{\sqrt{2\sigma^2}}$ وهذا هو متغير عشوائي طبيعي معياري أي أنّ: $\frac{X_1-X_2}{\sqrt{2\sigma^2}} \sim \mathrm{N}(0,1)$ ولو أخذنا مربّعه $\frac{(X_1-X_2)^2}{2\sigma^2} = \frac{\mathrm{Y}}{2\sigma^2}$ فنجد حسب النظرية أنّ $\frac{\mathrm{Y}}{2\sigma^2}$ ليس إلا متغيرا من النمط كاي مربع درجة حريته واحد وأنّ توقعه الرياضي يساوي درجة حريته وتباينه يساوي ضعف درجة حريته وبالتالي نجد أنّ:

$$X_1, X_2 \sim N(\mu,\sigma^2) \Rightarrow X_1 - X_2 \sim N(\mu,\sigma^2)$$
$$\Rightarrow \left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right) \sim N(0,1) \Rightarrow \left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$$

وبالمثل يكون لدينا:

$$\left(\frac{X_3-X_4}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)\ ,\ \left(\frac{X_5-X_6}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1), \left(\frac{X_7-X_8}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$$

$$\Rightarrow \frac{T_1}{a.2.\sigma^2} = \left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2 + \left(\frac{X_3-X_4}{\sqrt{2}\sigma}\right)^2 + \left(\frac{X_5-X_6}{\sqrt{2}\sigma}\right)^2 + \left(\frac{X_7-X_8}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(4)$$
$$\Rightarrow E\left(\frac{T_1}{a.2.\sigma^2}\right) = 4 \Rightarrow E(T_1) = (2\,a\,\sigma^2).(4) \Rightarrow E(T_1) = 8\,a\,\sigma^2$$

وحتى يكون T_1 غير منحاز بالنسبة لـ σ^2 يجب أنّ يكون :

$$E(T_1) = \sigma^2 \Rightarrow \sigma^2 = 8\,a\,\sigma^2 \Rightarrow a = \frac{1}{8}$$

$$\frac{T_1}{2a\sigma^2} = \frac{4T_1}{\sigma^2} \sim \chi^2(4) \Rightarrow Var\left(\frac{4T_1}{\sigma^2}\right) = (2)(4) \Rightarrow \frac{16}{\sigma^4} Var(T_1) = 8$$
$$Var(T_1) = \frac{\sigma^4}{2}$$
$$\frac{7T_2}{\sigma^2} = \sum_{i=1}^{8}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \sim \chi^2(7)$$
$$E\left(\frac{7T_2}{\sigma^2}\right) = 7 \Rightarrow \frac{7}{\sigma^2} E(T_2) = 7 \Rightarrow E(T_2) = \sigma^2$$

إذا T_2 غير منحاز لـ σ^2 .

$$Var\left(\frac{7T_2}{\sigma^2}\right) = 14 \Rightarrow \frac{49}{\sigma^4} Var(T_2) = 14 \Rightarrow Var(T_2) = \frac{14}{49}\sigma^4$$
$$Var(T_2) = \frac{2}{7}\sigma^4$$
$$Var(T_2) = \frac{2}{7}\sigma^4 < \frac{\sigma^4}{2} = Var(T_1)$$

إذا T_2 أفضل من T_1.

معلومات فيشر: لنفرض أنّه لدينا دالة كثافة $f(x,\theta)$ حيث $a \leq x \leq b$ و a,b محدودان ولنفرض أنّ $f(x,\theta)$ والمشتق الجزئي الأول لـ $f(x,\theta)$ بالنسبة إلى θ مستمرأنّ في المجال $[a,b]$ وأنّ المشتق الجزئي الثاني لـ $f(x,\theta)$ بالنسبة لـ θ موجود $\left(\frac{\partial^2 f(x,\theta)}{\partial\theta^2}\right)$ عندئذٍ يكون:

$$E\left[-\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right] = E\left[\frac{\partial \ln f(x,\theta)}{\partial\theta}\right]^2$$

البرهان:

بما أنّ $f(x,\theta)$ هي دالة كثافة احتمالية فإنّ: $\int_a^b f(x,\theta)dx = 1 \quad ....(1)$

وباشتقاق طرفي هذه العلاقة بالنسبة للوسيط θ فإنّنا نجد:

$$0 = \int_a^b \frac{\partial f(x,\theta)}{\partial\theta}dx = \int_a^b \frac{\partial f(x,\theta)}{\partial\theta}.\frac{f(x,\theta)}{f(x,\theta)}dx \Rightarrow \int_a^b \frac{\partial \ln f(x,\theta)}{\partial\theta}f(x,\theta)dx = 0 \qquad ...(2)$$

من العلاقة (2) نجد : $E\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right) = 0$

باشتقاق العلاقة (2) بالنسبة لـ θ نجد:

$$\int_a^b \left[\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}f(x,\theta)dx + \frac{\partial \ln f(x,\theta)}{\partial\theta}\frac{\partial \ln f(x,\theta)}{\partial\theta}f(x,\theta)\right]dx = 0$$

$$\Rightarrow \int_a^b \left[\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}f(x,\theta)\right]dx + \int_a^b \left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)dx = 0$$

$$E\left(-\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right) = E\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^2 = Var\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)$$

يسمي فيشر الكمية: $E\left(-\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right) = I$ بالمعلومات حول الوسيط θ التي يقدمها عنصر واحد من عناصر عينة عشوائية مأخوذة من دالة الكثافة $f(x,\theta)$.

تعميم: إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي دالة كثافته الاحتمالية $f(x,\theta)$ وبفرض أنّ $L(\underline{x},\theta) = L(x_1,\dots,x_n,\theta) = \prod_{i=1}^n f(x_i,\theta)$ هي كثافة العينة , فإنّ معلومات فيشر تعطى بالعلاقة التالية:

$$E\left(-\frac{\partial^2 \ln l(\underline{x},\theta)}{\partial\theta^2}\right) = E\left(-\sum_{i=1}^n \frac{\partial^2 \ln f(x_i,\theta)}{\partial\theta^2}\right) = nE\left(-\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right) = nI$$

أو بالعلاقة التالية:

$$Var\left(\frac{\partial \ln l(\underline{x},\theta)}{\partial\theta}\right) = Var\left(\sum_{i=1}^n \frac{\partial \ln f(x_i,\theta)}{\partial\theta}\right) = \sum_{i=1}^n Var\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right) = \sum_{i=1}^n I = nI$$

ونعبر عن ذلك بقولنا أنّ المعلومات حول الوسيط θ كما عرفها فيشر تتصف بخاصة التجميعية , وأنّ المعلومات حول θ في عينة تساوي حجم العينة مضروباً بالمعلومات حول θ في مشاهدة واحدة من مشاهداتها.

وإذا لم تحو الدالة $f(x,\theta)$ الوسيط θ فإنّ: $L(\underline{x},\theta) = 0$ أي أنّ المعلومات التي تقدمها عينة عشوائية من دالة كثافة لا تحوي θ حول الوسيط θ هي صفر كما هو متوقع.

مثال: أوجد معلومات فيشر للدالة $f(x,\theta)$ المعرفة بالشكل $f(x,\theta)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(x-\theta)^2}{2}}$

الحل: لدينا $E\left(-\frac{\partial^2 \ln f(\underline{x},\theta)}{\partial\theta^2}\right)=I$

نبدأ بإيجاد اللوغاريتم: $\ln f(x,\theta)=-\frac{1}{2}\ln(2\pi)-\frac{1}{2}(x-\theta)^2$

نشتق بالنسبة لـ θ : $\frac{\partial \ln f(x,\theta)}{\partial\theta}=0+(\mathrm{x}-\theta)$

نشتق مرة أخرى:

$$\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}=-1 \Rightarrow I=E\left(-\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right)=E(1)=1=Var(X-\theta)$$

متباينة كرامير– راو : لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي, ولتكن $L(\underline{x},\theta)$ هي كثافة العينة (الكثافة المشتركة لمتغيرات هذه العينة) وليكن: $T=T(X_1,X_2,\dots,X_n)$ تقديراً غير منحاز للوسيط $\varphi(\theta)$ أي $E(T)=\varphi(\theta)$ عندئذٍ يكون:

$$Var(T)\ge\frac{\left[\varphi'(\theta)\right]^2}{E\left(-\frac{\partial^2 \ln L(\underline{x},\theta)}{\partial\theta^2}\right)}$$

البرهان: بما أنّ T هو تقدير غير منحاز لـ $\varphi(\theta)$ فإنّ $E(T)=\varphi(\theta)$ لكن:

$$E(T)=\int_{R^n} tL(\underline{x},\theta)d\underline{x}=\varphi(\theta)$$

أي أنّ :

$$\int_{R^n} tL(\underline{x},\theta)d\underline{x}=\varphi(\theta)$$

وباشتقاق العلاقة الأخيرة جزئياً بالنسبة لـ θ نجد :

$$\int_{R^n} t\frac{\partial L(\underline{x},\theta)}{\partial\theta}d\underline{x}=\varphi'(\theta)$$

أي أنّ :

$$\varphi'(\theta)=\int_{R^n} t\frac{\partial L(\underline{x},\theta)}{\partial\theta}\frac{L(\underline{x},\theta)}{L(\underline{x},\theta)}d\underline{x}=\int_{R^n} t\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}L(\underline{x},\theta)d\underline{x}$$

وبالتالي نجد :

$$E\left(T\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)=\varphi'(\theta)$$

وبما أنّ التوقع الرياضي للمشتق الجزئي الأول للدالة $\ln L(\underline{x},\theta)$ يساوي الصفر أي: $E\left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)=0$ كما وجدنا سابقاً فإنّ:

$$E\left(T\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)=Cov\left(T,\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)$$

ولدينا معامل الارتباط بين متغيرين لا يتجاوز الواحد فإنّ:

$$Cov\left(T,\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)\leq\sqrt{VarT}.\sqrt{Var\left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)}$$

إذاً :

$$[\varphi'(\theta)]^2\leq Var(T).Var\left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)$$

وبتقسيم طرفي العلاقة على المقدار الثاني الموجب من الطرف الأيمن نجد:

$$Var(T)\geq\frac{[\varphi'(\theta)]^2}{Var\left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)}=\frac{[\varphi'(\theta)]^2}{E\left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)^2}=\frac{[\varphi'(\theta)]^2}{E\left(-\frac{\partial^2 \ln L(\underline{x},\theta)}{\partial\theta^2}\right)}$$

يسمى الطرف الأيمن بالحد الأدنى لتباين تقدير $\varphi(\theta)$ أو حد المعلومات لتباين تقدير $\varphi(\theta)$ وتقوم المساواة فقط عندما يكون $T-\varphi(\theta)$ متناسباً مع الكمية $\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}$ وذلك وفقاً لخواص متباينة كوشي – شوارتز.

الحد المنسوب إلى كرامير – راو *(Cramer – Raw)*

نظرية : لتكن لدينا $X_1,X_2,\dots,X_n$ عينة عشوائية ذات كثافة $f(x,\theta)$ حيث $\theta\in\Theta\in R$ وليكن $U=U(X_1,\dots,X_n)$ مقدِّر غير منحاز للوسيط θ فعندئذٍ يكون :

$$\mathrm{Var}(u)\geq\frac{1}{n\,E\left[\left(\frac{\partial\,\mathrm{Ln}\,f(x,\theta)}{\partial\theta}\right)^2\right]}$$

البرهان : سنفترض أنّ $f(x,\theta)$ تحقق شروط نظامية مثل قابلية الاشتقاق الجزئي بالنسبة للوسيط θ ، قابلية المكاملة ، قابلية تبديل إشارة التكامل مع المشتق الجزئي .

من جهة أولى نلاحظ أنّ : $1=\int_R f(x,\theta)\,dx$

وباشتقاق الطرفين جزئياً بالنسبة للوسيط θ فنجد أنّ :

$$0=\frac{\partial}{\partial\theta}\int_R f(x,\theta).\,dx=\int_R\frac{\partial f(x,\theta)}{\partial\theta}\,dx=\int_R\frac{\partial}{\partial\theta}f(x,\theta)\frac{1}{f(x,\theta)}f(x,\theta).\,dx$$

$$=\int_R\frac{\partial\,\mathrm{Ln}\,f(x,\theta)}{\partial\theta}f(x,\theta).\,dx=E\left[\frac{\partial\,\mathrm{Ln}\,f(x,\theta)}{\partial\theta}\right]$$

وبالتالي فإنّ : $(\,*\,)$

$$E\left[\frac{\partial\,\mathrm{Ln}\,f(x,\theta)}{\partial\theta}\right]=0$$

من جهة ثانية نجد أنّ $E(u)=\theta$ [u مقدِّر غير منحاز للوسيط θ] .

دالة الكثافة المشتركة : $\theta=E(u)=\int_{R^n}u\;g(u,\theta)\,du$

$$=\int_R\dots\int_R u(x_1,x_2,\dots,x_n)\,f(x_1,\theta)\dots f(x_n,\theta)\,dx_1\dots dx_n$$

وباشتقاق طرفي المساواة بالنسبة للوسيط θ سنجد أنّ :

$$1=\int_R\dots\int_R u(x_1,x_2,\dots,x_n)\frac{\partial}{\partial\theta}[f(x_1,\theta)\dots f(x_n,\theta)]dx_1\dots dx_n$$

$$=\int_R\dots\int_R u(x_1,\dots,x_n)\left[\sum_{i=1}^{n}\left(\frac{\partial f(x_i,\theta)}{\partial\theta}\prod_{j\neq i}f(x_j,\theta)\right)\right]dx_1\dots dx_n$$

$$=\int_R\dots\int_R u(x_1,\dots,x_n)\left[\sum_{i=1}^{n}\frac{\partial f(x_i,\theta)}{\partial\theta}\frac{f(x_i,\theta)}{f(x_i,\theta)}\prod_{j\neq i}f(x_j,\theta)\right]dx_1\dots dx_n$$

$$=\int_R\dots\int_R u(x_1,\dots,x_n)\left[\sum_{i=1}^{n}\left(\frac{\partial \operatorname{Ln} f(x_i,\theta)}{\partial\theta}\right)\right]\prod_{j\neq i}f(x_j,\theta)\,dx_1\dots dx_n$$

$$=\int_R\dots\int_R u(x_1,x_2,\dots,x_n)\,Z(x_1,x_2,\dots,x_n)\prod_{j=1}^{n}f(x_j,\theta)dx_1\dots dx_n$$

$$=E[U.Z]\Rightarrow E[U.Z]=1$$

ولندرس توقع المتحول العشوائي Z فنجد أنّ :

$$E[Z]=E\left[\sum_{i=1}^{n}\frac{\partial \operatorname{Ln} f(X_i,\theta)}{\partial\theta}\right]$$

$$=\sum_{i=1}^{n}E\left[\frac{\partial \operatorname{Ln} f(X_i,\theta)}{\partial\theta}\right]=0$$ (وذلك اعتماداً على *)

ولندرس تباين Z فنجد أنّ :

$$Var(Z)=\operatorname{Var}\left(\sum_{i=1}^{n}\frac{\partial \operatorname{Ln} f(X_i,\theta)}{\partial\theta}\right)$$

وذلك لأنّ X_i مستقلة :

$$=\sum_{i=1}^{n}\operatorname{Var}\left[\frac{\partial \operatorname{Ln} f(X_i,\theta)}{\partial\theta}\right]$$

وذلك لأنّ توقع هذا المتحول معدوم

$$=\sum_{i=1}^{n}E\left[\left(\frac{\partial \operatorname{Ln} f(X_i,\theta)}{\partial\theta}\right)^2\right]$$

$$=n\,E\left[\left(\frac{\partial \operatorname{Ln} f(X,\theta)}{\partial\theta}\right)^2\right]$$

وذلك لأنّ المتغيرات العشوائية مستقلة ولها نفس الكثافة :

$$\Rightarrow Var(Z)=n\,E\left[\left(\frac{\partial \operatorname{Ln} f(X,\theta)}{\partial\theta}\right)^2\right]$$

من جهة ثانية وجدنا أنّ : $E[U.Z]=1$, وكذلك نعلم حسب تعريف التغاير بين متحولين أنّ :

$$Cov(U,Z)=E[U.Z]-E(U)E(Z)$$

$$\Rightarrow \ \rho(U,Z)\sqrt{Var(U)}\ \sqrt{Var(Z)} = 1-0$$

$$\Rightarrow \qquad \rho^2(U,Z) = \frac{1}{Var(U)\ Var(Z)} \leq 1$$

$$\Rightarrow \quad Var(U) \geq \frac{1}{Var(Z)} \Rightarrow \ Var(U) \geq \frac{1}{n\,E\left[\left(\frac{\partial \, Ln\, f(X,\theta)}{\partial \theta}\right)^2\right]}$$

مثال : لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي يتبع التوزيع البواسوني ، قانونه الاحتمالي معرف بالشكل : $f(x,\theta) = e^{-\theta}\frac{\theta^x}{x!}$ ، حيث $x = 0\,,1\,,2\,,3\,,\ldots$ ، θ وسيط مجهول .

"1 – وسنبحث عن حد كرامير – راو لمقدِّر غير منحاز للوسيط θ .

$$Ln\, f(x,\theta) = x\,Ln\,\theta - \theta - Ln(x!)$$

$$\Rightarrow \frac{\partial\, Ln\, f(x,\theta)}{\partial \theta} = \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}$$

ومنه فإنّ :

$$E\left[\frac{\partial}{\partial \theta} Ln\, f(x,\theta)\right] = E\left[\left(\frac{X-\theta}{\theta}\right)^2\right]$$

$$E\left[\left(\frac{\partial}{\partial \theta} Ln\, f(x,\theta)\right)^2\right] = \frac{1}{\theta^2}E(X-\theta)^2$$

$$= \frac{1}{\theta^2}\ Var(X) \ = \frac{1}{\theta}$$

(وذلك لأنّ $Var(X) = \theta$)

وبالتالي سنجد أنّ :

$$B.C.R = \frac{1}{n\left(\frac{1}{\theta}\right)} = \frac{\theta}{n}$$

"2 – لنقدِّر الوسيط θ بواسطة المقدِّر $\overline{X}$ (المتوسط) ولنبحث في جودة هذا المقدِّر : ولذلك سنحسب كل من توقعه الرياضي وتباينه أيضاً فنجد أنّ :

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta$$

ومنه فإنّ $\overline{X}$ مقدِّر غير منحاز .

ونجد أنّ تباينه سيكتب بالشكل التالي :

$$Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}Var(X) = \frac{\theta}{n}$$

ومنه فإنّ تباين $\overline{X}$ يساوي الحد الأدنى لكرامير – راو وبالتالي فإنّ $\overline{X}$ مقدِّر غير منحاز ذو تباين أصغري .

مثال: لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه $\mu = 0\,,\ \sigma^2 = \theta$. استخدم متباينة كرامير – راو في إثبات أنّ التقدير $T = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ يبلغ الحد الأدنى للتباين .

الحل:

لدينا دالة الكثافة للتوزيع الطبيعي معرفة بالشكل:

$$f(x,\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{1}{2}\frac{x^2}{\theta}}$$

وبأخذ اللوغاريتم الطبيعي للطرفين نجد:

$$Lnf(x,\theta)=-\frac{1}{2}Ln(2\pi)-\frac{1}{2}Ln(\theta)-\frac{1}{2}\frac{x^2}{\theta}$$

وباشتقاق الطرفين جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial Lnf(x,\theta)}{\partial\theta}=-\frac{1}{2\theta}+\frac{x^2}{2\theta^2}$$

وبالاشتقاق مرة أخرى جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial^2 Lnf(x,\theta)}{\partial\theta^2}=\frac{1}{2\theta^2}-\frac{x^2}{\theta^3}$$

لنوجد الآن معلومات فيشر من العلاقة:

$$I=E\left(-\frac{\partial^2 Lnf(x,\theta)}{\partial\theta^2}\right)=E\left(\frac{X^2}{\theta^3}-\frac{1}{2\theta^2}\right)=\frac{1}{\theta^3}E(X^2)-E\left(\frac{1}{2\theta^2}\right)$$
$$=\frac{1}{\theta^3}Var(X)-\frac{1}{2\theta^2}=\frac{1}{\theta^3}(\theta)-\frac{1}{2\theta^2}=\frac{1}{\theta^2}-\frac{1}{2\theta^2}=\frac{1}{2\theta^2}$$

وبالتالي فإنّ الحد الأدنى للتباين كما تحدده متباينة كرامير – راو يساوي $\frac{2\theta^2}{n}$.

لنوجد الآن تباين المقدّر $T=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ بالشكل:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i^2\right)$$
$$=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i^2)=\frac{1}{n^2}\left[n\left(Var(X^2)\right)\right]=\frac{1}{n}Var(X^2)$$

وذلك لأنّ المتغيرات العشوائية $X_1,X_2,\dots,X_n$ مستقلة , ولكن تباين المتغير X^2 يحسب من العلاقة التالية:

$$Var(X^2)=E(X^4)-[E(X^2)]^2=(\mu_4-\mu_2^2)$$
$$=\left[\left(\frac{4!}{(2^2)(2!)}\right)\theta^2-\theta^2\right]=3\theta^2-\theta^2=2\theta^2$$

علماً أنّ عزم المجتمع من المرتبة الرابعة حول الصفر للمتغير الطبيعي يعطى من العلاقة التالية:

$\mu_{2r}=\frac{(2r)!}{2^r(r!)}\theta^2$, حيث لدينا في هذا المثال أنّ : $\sigma^2=\theta$.

وبالتالي فإنّ:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{2\theta^2}{n}$$

بالموازنة بين تباين المقدّر والحد الأدنى للتباين كما تحدده متباينة كرامير – راو نجد أنّهما متساويان وبالتالي فإنّ هذا المقدّر هو مقدّر غير منحاز وذو تباين أصغري.

طرائق التقدير: هناك عدة طرائق لإيجاد تقدير ما هي:

أولاً: مبدأ الاحتمالية العظمى:

تعريف: أنّ دالة الاحتمالية لـ n متغيراً عشوائياً $X_1, X_2, \dots, X_n$, هي دالة الكثافة المشتركة لهذه المتغيرات العشوائية ونرمز لها $L(\underset{\sim}{x}, \theta)$ وهي تحوي بشكل عام وسيط θ أو أكثر.

وبشكل خاص إذا شكلت المتغيرات العشوائية $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من المجتمع الإحصائي $f(x, \theta)$ فإنّ دالة الاحتمالية $L(\underset{\sim}{x}, \theta)$ هي: $L(\underset{\sim}{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$.

وتعبر دالة الاحتمالية بشكل عام عن إمكانية أنّ تأخذ المتغيرات العشوائية قيمة معينة $(x_1, x_2, \dots, x_n)$. فإذا فرضنا أنّ θ معلومة ولتكن θ_0 فإنّ القيمة الخاصة للمتغيرات العشوائية التي يكون وقوعها (تحققها) هو الأكثر احتمالاً هي القيمة: $(\acute{x}_1, \acute{x}_2, \dots, \acute{x}_n)$ التي تجعل دالة الاحتمالية $L(\underset{\sim}{x}, \theta_0)$ أكبر ما يمكن, وعندما تأخذ θ قيمها المختلفة في فضاء الوسيط Θ فإنّها تعرف من أجل كل قيمة , دالة كثافة معينة وبالتالي مجتمعاً معيناً.

وبعد أنّ نحصل على عينة إحصائية $(\acute{x}_1, \acute{x}_2, \dots, \acute{x}_n)$ للمتغيرات $(x_1, x_2, \dots, x_n)$ فإنّنا نريد أنّ نتعرف على ذلك المجتمع الذي يمكن أنّ يعطي مثل هذه العينة بأكبر احتمال ممكن.

وبعبارة أخرى: نريد إيجاد قيمة لـ θ في الفضاء Θ نرمز لها $\hat{\theta}$ تجعل دالة الاحتمالية $L(\underset{\sim}{x}, \theta)$ في نهايتها العظمى, وقيمة $\hat{\theta}$ هذه ستكون بشكل عام تابعاً في $(x_1, x_2, \dots, x_n)$ أي أنّها متغير عشوائي أو إحصاء نسميه عادةً بتقدير الاحتمالية العظمى لـ θ .

1 – شرح الطريقة من أجل وسيط واحد:

1) نشكل دالة الاحتمالية للمتغيرات العشوائية $X_1, X_2, \dots, X_n$ نرمز لها $L(\underset{\sim}{x}, \theta)$.

2) نشتق دالة الاحتمالية بالنسبة للوسيط θ ونبدل كل θ بـ $\hat{\theta}$.

3) نحل جملة المعادلتين :

$$\left.\frac{\partial L(\underset{\sim}{x}, \theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0 \quad , \quad \left.\frac{\partial^2 L(\underset{\sim}{x}, \theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0$$

2 – شرح الطريقة من أجل k وسيط :

أنّ التقدير المشترك: $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ هو الحل المشترك لجملة المعادلات:

$$\left.\frac{\partial L(\underset{\sim}{x}, \theta)}{\partial \theta_1}\right|_{\theta_1=\hat{\theta}_1} = 0 \; , \; \left.\frac{\partial L(\underset{\sim}{x}, \theta)}{\partial \theta_2}\right|_{\theta_2=\hat{\theta}_2} = 0 \; , \dots , \; \left.\frac{\partial L(\underset{\sim}{x}, \theta)}{\partial \theta_k}\right|_{\theta_k=\hat{\theta}_k} = 0$$

$$\left.\frac{\partial^2 L(\underset{\sim}{x}, \theta)}{\partial \theta_1^2}\right|_{\theta_1=\hat{\theta}_1} < 0 \; , \; \left.\frac{\partial^2 L(\underset{\sim}{x}, \theta)}{\partial \theta_2^2}\right|_{\theta_2=\hat{\theta}_2} < 0 \; , \dots , \; \left.\frac{\partial^2 L(\underset{\sim}{x}, \theta)}{\partial \theta_k^2}\right|_{\theta_k=\hat{\theta}_k} < 0$$

ملاحظة هامة: بما أنّ L و $ln\,L$ تبلغان نهايتيهما العظمى عند نفس النقطة فيكون أسهل أحياناً أنّ نحسب النهاية العظمى للوغاريتم دالة الاحتمالية بدلاً من النهاية العظمى لدالة الاحتمالية , ففي حالة وسيط واحد نحل جملة المعادلتين:

$$\left.\frac{\partial lnL(\underline{x},\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}=0\quad,\quad\left.\frac{\partial^2 lnL(\underline{x},\theta)}{\partial\theta^2}\right|_{\theta=\hat{\theta}}<0$$

أما في حالة k وسيط فإنّ التقدير المشترك: $\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_{\mathrm{k}}$ هو الحل المشترك لجملة المعادلات.

$$\left.\frac{\partial lnL(\underline{x},\underline{\theta})}{\partial\theta_1}\right|_{\theta_1=\hat{\theta}_1}=0\quad,\quad\left.\frac{\partial lnL(\underline{x},\underline{\theta})}{\partial\theta_2}\right|_{\theta_2=\hat{\theta}_2}=0\;,\dots,\;\left.\frac{\partial lnL(\underline{x},\underline{\theta})}{\partial\theta_{\mathrm{k}}}\right|_{\theta_k=\hat{\theta}_k}=0$$

$$\left.\frac{\partial^2 lnL(\underline{x},\theta)}{\partial\theta_1^2}\right|_{\theta_1=\hat{\theta}_1}<0\quad,\quad\left.\frac{\partial^2 lnL(\underline{x},\theta)}{\partial\theta_2^2}\right|_{\theta_2=\hat{\theta}_2}<0\;,\dots,\;\left.\frac{\partial^2 lnL(\underline{x},\theta)}{\partial\theta_k^2}\right|_{\theta_k=\hat{\theta}_k}<0$$

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه (μ,σ^2) ودالة كثافته : $f(x,\theta)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

والمطلوب قدر كل من الوسيطين μ , σ^2 باستخدام مبدأ الاحتمالية العظمى.

الحل: عدد الوسطاء 2 لذلك نريد إيجاد معادلتين.

1) نشكل دالة الاحتمالية العظمى: $L(\underline{x},\mu,\sigma^2)=(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}e^{-\sum(x_i-\mu)^2/2\sigma^2}$

2) نأخذ اللوغاريتم الطبيعي للدالة الاحتمالية:

$$ln\,L(\underline{x},\mu,\sigma^2)=-\frac{n}{2}ln(2\pi)-\frac{n}{2}ln\,\sigma^2-\frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

3) نشتق جزئياً بالنسبة لـ μ ثم بالنسبة لـ σ^2 :

$$\left.\frac{\partial\;ln\,L(\underline{x},\mu,\sigma^2)}{\partial\,\mu}\right|_{\mu=\hat{\mu}}=\left.\frac{2\sum(x_i-\mu)}{2\sigma^2}\right|_{\mu=\hat{\mu}}=\frac{\sum(x_i-\hat{\mu})}{\sigma^2}=0\Longrightarrow$$

$$\sum(x_i-\hat{\mu})=0\Longrightarrow\sum x_i-\hat{\mu}=0\Longrightarrow\hat{\mu}=\frac{\sum x_i}{n}=\overline{X}$$

$$\left.\frac{\partial\;ln\,L(\underline{x},\mu,\sigma^2)}{\partial\,\mu}\right|_{\sigma^2=\hat{\sigma}^2}=-\frac{n}{\sigma^2}+\left.\frac{\sum(x_i-\overline{x})^2}{2\sigma^2}\right|_{\sigma^2=\hat{\sigma}^2}=0\Longrightarrow$$

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=n\hat{\sigma}^2\Longrightarrow\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$$

خواص تقدير الاحتمالية العظمى من أجل عينات كبيرة الحجم: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي قانونه $f(x,\theta)$ حيث θ هو الوسيط, وليكن $\hat{\theta}$ هو تقدير الاحتمالية العظمى لـ θ عندئذٍ يكون:

1) التقدير $\hat{\theta}$ متّسق, أي من أجل $0<\varepsilon$ يكون $p(|\hat{\theta}-\theta|>\varepsilon)\xrightarrow[n\to\infty]{}0$ أو أنّ $\hat{\theta}$ يتقارب احتمالياً إلى θ.

2) يتقارب توزيع التقدير $\hat{\theta}$ إلى التوزيع الطبيعي, أي أنّ $\sqrt{n}\left(\hat{\theta}-\theta\right)\xrightarrow[n\to\infty]{}N(0,c^2)$ حيث c^2 هي معكوس مصفوفة المعلومات لعينة, أي أنّ التباين يدرك أحد المعلومات عندما تسعى n إلى اللانهاية.

3) يتقارب التقدير $\hat{\theta}$ إلى أنّ يكون التقدير الفعال, أي أنّ أي تقدير آخر ينتهي إلى التوزيع الطبيعي, سيكون له تباين مقارب أكبر أو يساوي c^2 ويمكن تعميم ذلك إلى أكثر من وسيط.

ثانياً : طريقة العزوم .

لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي الذي يكون تابعاً بشكلٍ عام لـ k وسيط, أي مأخوذة من المجتمع $f(x,\theta_1,\dots,\theta_k)$, لإيجاد مقدِّرات لهذه الوسطاء الإحصائية والتي نرمز لها بالرموز $\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k$ بطريقة العزوم:

نحسب أولاً التوقع الرياضي للمتغيرات العشوائية $X,X^2,\dots,X^k$ حيث k عدد الوسطاء, وهذا يكافئ أنّ نوجد العزوم حتى المرتبة k حول الصف للمتغير العشوائي X وتكون هذه التوقعات الرياضية أو هذه العزوم تابعة بشكل عام إلى الوسطاء الإحصائية $\theta_1,\theta_2,\dots,\theta_k$, أي نكتب:

$$\left.\begin{array}{r} E(X)=g_1(\theta_1,\theta_2,\dots,\theta_k) \\ E(X^2)=g_2(\theta_1,\theta_2,\dots,\theta_k) \\ \dots \\ E(X^k)=g_k(\theta_1,\theta_2,\dots,\theta_k) \end{array}\right\} \quad (1)$$

عدد هذه الوسطاء يساوي عدد المعادلات k , ثم نختار ثانياً مقدِّرات لهذه الوسطاء وهي: $\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k$ ثم نضع:

$$\left.\begin{array}{r} g_1\left(\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k\right)=\frac{1}{n}\sum_{i=1}^{n}X_i \\ g_2\left(\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k\right)=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \\ \dots \\ g_k\left(\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k\right)=\frac{1}{n}\sum_{i=1}^{n}X_i^k \end{array}\right\} \quad (2)$$

نحل جملة المعادلات (2) حلاً مشتركاً فنحصل على المقدِّرات $\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_k$ للوسطاء $\theta_1,\theta_2,\dots,\theta_k$.

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من المجتمع الإحصائي الذي له التوزيع المنتظم, بين الصفر و θ. والمطلوب, قدر الوسيط θ باستخدام طريقة العزوم ثم بطريقة الاحتمالية العظمى.

الحل:

المجتمع الإحصائي هو: $f(x,\theta)=\frac{1}{\theta}\ ,\ 0\le x\le\theta \Leftrightarrow x\in\ [0,\theta]$

أولاً – طريقة العزوم:

1– لنوجد التوقع الرياضي لـ X :

$$E(X) = \int_{-\infty}^{+\infty} xf(x).dx = \int_0^{\theta} x\frac{1}{\theta}dx = \frac{x^2}{2\theta}\Big|_0^{\theta} = \frac{\theta}{2}$$

2− نبدل كل θ بـ $\hat{\theta}$ ونكتب: $\frac{\hat{\theta}}{2} = \frac{1}{n}\sum X_i = \overline{X}$

نحل المعادلة فنجد: $\hat{\theta} = 2\overline{X}$.

ثانياً – طريقة الاحتمالية العظمى:

لدينا دالة الكثافة الاحتمالية :

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n}\left(\frac{1}{\theta}\right) = \left(\frac{1}{\theta}\right)^n$$

حيث أن $0 \le x \le \theta$.

نأخذ لوغاريتم الطرفين فنجد :

$$lnL(x_1, x_2, \ldots, x_n; \theta) = -nln\theta$$

$$\frac{\partial\, lnL(x_1,x_2,\ldots,x_n;\theta)}{\partial\theta}\Big|_{\mu=\hat{\mu}} = -\frac{n}{\theta}\Big|_{\theta=\hat{\theta}} = 0 \Longrightarrow \hat{\theta} = \infty$$

وهذا غير منطقي ، أي أننا لا نستطيع إيجاد المقدر بهذا الشكل لذلك سوف نجد مقدر الاحتمالية العظمى لهذا الوسيط بشكل آخر .

وبملاحظة أن دالة الكثافة $L(x_1, x_2, \ldots, x_n; \theta)$ تكون أعظم ما يمكن عندما تكون θ أصغر ما يمكن , أي نجد أنها متناقصة في θ وبالتالي . وبما أن $X_i \le \theta \;\; ; \;\; i = 1,2, \ldots, n$ عندئذٍ تكون أصغر قيمة للوسيط θ هي أعظم قيمة مرتبة لمتغيرات العينة العشوائية أي هي $Z = max\{X_1, X_2, \ldots, X_n\}$ وهذا يعني أن مقدر الاحتمالية العظمى هو :

$$\hat{\theta} = Z = max\{X_1, X_2, \ldots, X_n\}$$

وقد بيّنا في فقرة سابقة أن دالة الكثافة الاحتمالية للمتغير Z هي :

$$f(z, \theta) = \frac{n}{\theta^n} z^{n-1} \;\; ; \;\; 0 \le z \le \theta$$

ووجدنا أيضاً أن :

$$E(Z) = E(\hat{\theta}) = \left(\frac{n}{n+1}\right)\theta$$

أي أن مقدر الاحتمالية العظمى للوسيط θ هو مقدر منحاز.

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع طبيعي وسيطاه μ, σ^2 والمطلوب:

استخدم طريقة العزوم في تقدير كل من الوسيطين μ, σ^2 .

الحل: نعلم أنّ $E(X) = \mu \quad , \quad E(X^2) = \sigma^2 + \mu^2$

نستبدل كل μ بـ $\hat{\mu}$ و كل σ^2 بـ $\widehat{\sigma^2}$ فنجد:

$$\hat{\mu} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} X_i$$

$$\hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} X_i^2$$

نحل جملة هاتين المعادلتين ونوجد التقديرين $\hat{\sigma}^2$ و $\hat{\mu}$.

من المعادلة الأولى نجد: $\hat{\mu} = \bar{X}$

من المعادلة الثانية نجد: $\hat{\sigma}^2 = \frac{1}{n} \sum X_i^2 - (\bar{X})^2 \Longrightarrow \hat{\sigma}^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (X_i - \bar{X})^2$

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع طبيعي وسيطاه μ, σ^2 والمطلوب:

1 – أثبت أنّ المقدِّر $\hat{\mu} = \bar{X}$ هو مقدِّر غير منحاز بالنسبة للوسيط μ , ثم أحسب تباينه.

2 – بين فيما إذا المقدِّر $\hat{\sigma}^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (X_i - \bar{X})^2$ غير منحاز بالنسبة للوسيط σ^2 .

الحل:

1 – حتى يكون المقدِّر $\hat{\mu}$ غير منحاز بالنسبة للوسيط μ يجب أنّ يكون: $E(\hat{\mu}) = \mu$.

$$E(\hat{\mu}) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

وبالتالي فإنّ المقدِّر $\hat{\mu}$ هو مقدِّر غير منحاز بالنسبة للوسيط μ .

لنوجد الآن تباين المقدِّر $\hat{\mu}$:

$$Var(\hat{\mu}) = Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

وذلك بالاستفادة من خواص التباين والاستقلال العشوائي.

2 – سوف نحاول كتابة

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{2}{n} \sum_{i=1}^{n} (X_i - \mu)(\bar{X} - \mu) + \frac{1}{n} \sum_{i=1}^{n} (\bar{X} - \mu)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{2}{n} (\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) + \frac{1}{n} [n(\bar{X} - \mu)^2] \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{2}{n} (\bar{X} - \mu)[n(\bar{X} - \mu)] + n(\bar{X} - \mu)^2
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\bar{X}-\mu)^2 + (\bar{X}-\mu)^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\bar{X}-\mu)^2$$

لنوجد الآن التوقع الرياضي:

$$E(\hat{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\bar{X}-\mu)^2\right]$$
$$= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] - E[(\bar{X}-\mu)^2]$$
$$= \frac{1}{n}\sum_{i=1}^{n}E(X_i-\mu)^2 - E[(\bar{X}-\mu)^2]$$
$$= \frac{1}{n}\sum_{i=1}^{n}Var(X_i) - Var(\bar{X})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2}{n} = \frac{1}{n}[n\sigma^2] - \frac{\sigma^2}{n}$$
$$= \sigma^2 - \frac{\sigma^2}{n} = \left(1-\frac{1}{n}\right)\sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

مما يدل على أنّ هذا المقدِّر منحاز .

وبقسمة الطرفين على المقدار $\frac{n-1}{n}$ نجد:

$$\left(\frac{n}{n-1}\right)E(\hat{\sigma}^2) = \sigma^2 \Rightarrow E\left[\left(\frac{n}{n-1}\right)\hat{\sigma}^2\right] = \sigma^2$$

وإذا رمزنا للكمية $\left(\frac{n}{n-1}\right)\hat{\sigma}^2$ بـ S^2 حيث $S^2 = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{n}}(X_i-\bar{X})^2$ فإنّ هذه الكمية S^2 تصبح مقدِّر غير منحاز للوسيط σ^2 .

ثالثاً : تقديرات بايز .

في الفقرات السابقة كنا نعالج مسألة تقدير الوسيط θ في تابع الكثافة $f(x,\theta)$ على أساس أنّ θ وسيط مجهول يأخذ قيمه في فراغ Ω يسمى فراغ الوسيط . أي أنّه تبرز أحياناً حالات تتوفر فيها معلومات إضافية حول θ . فيمكن أنّ تتوفر لدى المجرب مثلاً دلائل على أنّ θ تتغير وكأنها هي الأخرى متحول عشوائي وأنّه يتمكن عملياً من افتراض تابع كثافة معين لـ θ .

فلنفرض أنّ آلة تنتج نوعاً ما من قطع الغيار , وأنّ هذه القطع يجب أنّ تخضع لفحص يؤكد صلاحيتها للاستعمال أو عدم صلاحيتها, ويحدد نسبة القطع الغير الصالحة من الإنتاج اليومي لهذه الآلة. وفي يوم معين جرى اختبار عشر قطع, ولنرمز لنتائج الاختبار بـ $X_1, X_2, \dots, X_{10}$ حيث $X_i = 1$ إذا كانت القطعة i غير صالحة للاستعمال , و $X_i = 0$ إذا كانت صالحة للاستعمال.

فيمكن النظر إلى نتيجة هذا الاختبار وكانه عينة عشوائية من التوزيع الهندسي:

$$f(x,p) = p^x(1-p)^{1-x} \qquad , \qquad x = 0,1 \quad , \quad 0 \le p \le 1$$

وتابع الكثافة المشترك لمتحولات العينة هو:

$$L(x_1,x_2,\dots,x_{10},p)=p^{\sum_{i=1}^{10}x_i}q^{10-\sum_{i=1}^{10}x_i} \quad , \quad x_i=0{,}1 \quad , \quad 0\le p\le 1$$

وتقدير الاحتمالية العظمى في هذه الحالة أي $\hat{p}=\overline{X}$ هو التقدير الذي يجب أنّ نحصل عليه عند استخدام طريقة العزوم أيضاً. ولنفرض الآن أنّ المجرب قد توفرت لديه معلومات إضافية يبدو له معها أنّ النسبة p تتغير من يوم لآخر, وأنّ تغيرها هذا يمكن تمثيله بتغير متحول عشوائي تابع كثافته $h(p)=6(1-p)$ حيث $0\le p\le 1$, والسؤال الآن كيف يمكن استخدام مثل هذه المعلومات الإضافية في مسألة تقدير p ؟

و الواقع أنّه في العديد من مثل هذه الأسئلة لا يبدو واقعياً أو منطقياً أنّ نفترض أنّ p متحول عشوائي , وبالرغم من أنّه يبدو مقبولاً أحياناً أنّ نفترض أنّ p تتغير وكأنها متحول عشوائي إلا أنّ تابع كثافة أو توزيع p يمكن أم لا يكون معروفاً لنا, وسنقتصر في الفقرة على الحالة التي يكون فيها الفرض بأنّ الوسيط متحول عشوائي وأنّ له تابع كثافة معين هو فرض واقعي و مقبول. وسنرمز لتابع الكثافة لمتحولات العينة بالرمز $f(x,\theta)$ للدلالة على أنّنا ننظر لـ θ وكأنها متحول عشوائي.

لنفرض الآن عينة عشوائية $X_1,X_2,\dots,X_n$ من تابع الكثافة $f(x|\theta)$ وليكن $f(\theta)$ تابع الكثافة الهامشي لـ θ و $I(\hat{\theta},\theta)$ تابع الخسارة, ولنتذكر أنّه بالرغم من أنّ θ متحول عشوائي إلا أنّنا نريد تقدير تلك القيمة لـ θ ولتكن $\hat{\theta}$ التي تحدد تابع الكثافة $f(x|\theta)$ وبالتالي المجتمع الذي أتت منه العينة الخاصة التي حصلنا عليها. وبعبارة أخرى فإنّ θ تحدد من أجل كل قيمة لها تابع كثافة $f(x|\theta)$ وبالتالي مجتمعاً ما. والعينة التي حصلنا عليها أتت من إحدى هذه المجتمعات والمطلوب تقدير قيمة θ الخاصة بهذا المجتمع بالذات . أنّ تابع المخاطرة هو كما نعلم : $R(d,\theta)=E\left[I(\hat{\theta},\theta)\right]$, وبما أنّ R تابعة لـ θ فقد وجدنا في مطلع هذا الفصل أنّ إحدى الطرق لاختيار أفضل تابع قرار d هو أنّ نحسب توقع R فوق قيم θ المختلفة ونتبنى التابع d الذي يجعل $E_\theta[R(d,\theta)]$ أصغر ما يمكن , ويمكن أنّ نكتب توقع المخاطرة على الشكل:

$$B(d)=E[R(d,\theta)]=\int_{-\infty}^{+\infty}R(d,\theta).f(\theta).d\theta= \qquad (1)$$

$$=\int_{-\infty}^{+\infty}\{\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}I[d(x_1,x_2,\dots,x_n),\theta]\,L(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n\}f(\theta)d\theta$$

وتقدير بايز هو التابع $\hat{\theta}=d(x_1,x_2,\dots,x_n)$ الذي يجعل $B(d)$ أصغر ما يمكن. وإذا غيرنا ترتيب التكامل في (1) بالنسبة لـθ والمتغيرات x_i نجد:

$$B(d)=\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}I[d(x_1,x_2,\dots,x_n),\theta]L(x_1,x_2,\dots,x_n|\theta)f(\theta)d\theta\right\}dx_1dx_2\dots dx_n \qquad (2)$$

والتابع d المطلوب هو التابع الذي يجعل الكمية بين قوسين في (2) أصغر ما يمكن , وذلك مهما كانت قيم المتغيرات x_i , أي أنّنا نريد التابع d الذي يعطي النهاية الصغرى للكمية:

$$\int_{-\infty}^{+\infty}I[d\theta]L(x_1,x_2,\dots,x_n|\theta)f(\theta)d\theta \qquad (3)$$

ولكن جداء $L(x_1,x_2,\dots,x_n|\theta)f(\theta)$ يمثل تابع الكثافة المشترك للمتحولات $X_1,X_2,\dots,X_n,\theta$ وسنرمز بـ $q(x_1,x_2,\dots,x_n;\theta)$ ، وتابع الكثافة الهامشي للمتحولات x_i هو:

$$h(x_1,x_2,\dots,x_n)=\int_{-\infty}^{+\infty}q(x_1,x_2,\dots,x_n;\theta)d\theta$$
$$=\int_{-\infty}^{+\infty}L(x_1,x_2,\dots,x_n;\theta\,)f(\theta)d\theta \dots \;(4)$$

والتوزيع الشرطي لـ θ علماً أنّ $X_1,X_2,\dots,X_n$ معروفة هو:

$$K(\theta|x_1,x_2,\dots,x_n)=\frac{q(x_1,x_2,\dots,x_n;\theta\,)}{\mathrm{h}(x_1,x_2,\dots,x_n)}=\frac{L(x_1,x_2,\dots,x_n|\theta)f(\theta)}{\mathrm{h}(x_1,x_2,\dots,x_n)}\quad \dots(5)$$

ويسمى هذا التوزيع بالتوزيع اللاحق لـ θ ، بينما نسمي التوزيع $f(\theta)$ بالتوزيع السابق أو التوزيع السلفي. وهكذا يمكن أنّ نكتب (3) على الشكل:

$$h(x_1,x_2,\dots,x_n)=\int_{-\infty}^{+\infty}I[d\theta]\,K(\theta|x_1,x_2,\dots,x_n)d\theta \quad \dots(6)$$

وتقدير بايز هو المقدار $d=\hat{\theta}$ الذي يجعل التابع $V(\hat{\theta},x_1,x_2,\dots,x_n)$ أصغر ما يمكن حيث:

$$V\big(\hat{\theta},x_1,x_2,\dots,x_n\big)=\int_{-\infty}^{+\infty}I\big(\hat{\theta},\theta\big)K\,(\theta|x_1,x_2,\dots,x_n)d\theta \quad \dots(7)$$

وذلك من أجل جميع القيم الممكنة للعينة $x_1,x_2,\dots,x_n$ ، ويمثل التابع V المخاطرة اللاحقة من أجل تقدير θ علماً أنّ $X_1=x_1,X_2=x_2,\dots,X_n=x_n$ حيث $x_1,x_2,\dots,x_n$ قيم العينة التي حصلنا عليها فعلاً.

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له توزيع برنولي المعرف بدالة الكثافة:

$$f(x|\theta)=\theta^{x}(1-\theta)^{1-x};x=0,1 \quad ; \quad 0<\theta<1$$

وبفرض أن التوزيع السلفي لـ θ هو التوزيع المنتظم على الفترة $[0,1]$:

$$\rho(\theta)=1 \quad ; \quad 0<\theta<1$$

وأن دالة الخسارة هي دالة الخطأ التربيعي $l\big(\hat{\theta},\theta\big)=(\hat{\theta}-\theta)^2$ ، والمطلوب:

1 – أوجد التوزيع اللاحق لـ θ .

2 – أوجد تقدير بايز لـ θ .

الحل:

1 – لدينا دالة الكثافة المشتركة لمتغيرات العينة:

$$f(x_1,x_2,\dots,x_n|\theta)=\theta^{\sum_{i=1}^{n}x_i}\,(1-\theta)^{n-\sum_{i=1}^{n}x_i}$$

عندئذٍ يكون :

$$q\,(x_1,x_2,\dots,x_n\,;\theta)=1\times\,\theta^{\sum_{i=1}^{n}x_i}\,(1-\theta)^{n-\sum_{i=1}^{n}x_i}=\theta^{\sum_{i=1}^{n}x_i}\,(1-\theta)^{n-\sum_{i=1}^{n}x_i}$$

وكذلك يكون:

$$h(x_1,x_2,\dots,x_n)=\int_0^1\theta^{\sum_{i=1}^{n}x_i}\,(1-\theta)^{n-\sum_{i=1}^{n}x_i}\,d\theta$$

$$= Be(y+1, n-y+1) = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} \;\; ; \; y = \sum_{i=1}^{n} x_i$$

$$= \frac{[(\sum_{i=1}^{n} x_i)!][(n-\sum_{i=1}^{n} x_i)!]}{(n+1)!}$$

وبالتالي فإن التوزيع اللاحق لـ θ يعطى بدالة الكثافة التالية:

$$K(\theta|x_1, x_2, \dots, x_n) = \frac{1}{Be(y+1, n-y+1)} \theta^{y}(1-\theta)^{n-y} \quad ; \;\; 0 < \theta < 1$$

$$= \frac{(n+1)! \;\; \theta^{\sum_{i=1}^{n} x_i} \;\; (1-\theta)^{n-\sum_{i=1}^{n} x_i}}{[(\sum_{i=1}^{n} x_i)!][(n-\sum_{i=1}^{n} x_i)!]} \qquad ; 0 \le \theta \le 1$$

2 – ويبقى المطلوب هو الدالة $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ التي تجعل توقع دالة الخطأ التربيعي V أصغر ما يمكن:

$$Var(\hat{\theta}, x_1, x_2, \dots, x_n) = \int_0^1 (\hat{\theta} - \theta)^2 \; \frac{(n+1)! \, \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}}{[(\sum_{i=0}^{n} x_i)!][(n-\sum_{i=0}^{n} x_i)!]} \; d\theta$$

$$= \hat{\theta}^2 - 2\hat{\theta} \frac{[(n-1)!][\sum_{i=0}^{n} x_i+1)!]}{[(n+2)!][\sum_{i=0}^{n} x_i)!]} + \frac{[(n+1)!][\sum_{i=0}^{n} x_i+2)!]}{[(n+3)!][\sum_{i=0}^{n} x_i)!]}$$

$$= \hat{\theta}^2 - 2\hat{\theta} \frac{\sum x_i + 1}{n+2} + \frac{(\sum x_i + 2)(\sum x_i + 1)}{n+3}$$

وقيمة $\hat{\theta}$ كتابع في المتغيرات x_i التي تجعل V أصغر ما يمكن هي : $\hat{\theta} = \frac{\sum x_i + 1}{n+2}$

أي أنّ تقدير بايز للوسيط θ هو: : $\hat{\theta} = \frac{\sum x_i + 1}{n+2}$.

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N(\theta, 1)$ والمعرف بدالة الكثافة:

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \;\; ; \; x \in R \;\; , \;\; \theta \in R$$

وبفرض أن التوزيع السلفي لـ θ هو التوزيع الطبيعي المعياري $N(0,1)$ والمعرف بدالة الكثافة:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} \;\; ; \; x \in R \;\; , \;\; \theta \in R$$

وأن دالة الخسارة هي دالة الخطأ التربيعي $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, والمطلوب:

1 – أوجد التوزيع اللاحق لـ θ .

2 – أوجد تقدير بايز لـ θ .

الحل:

لدينا دالة الكثافة المشتركة لمتغيرات العينة:

$$f(x_1,x_2,\dots,x_n|\theta)=\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}=\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\left(\sum_{i=1}^{n}x_i{}^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right)}$$

ولنفرض أنّ μ متحول عشوائي تابع كثافته: $p(\mu)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\mu^2}\quad,\ -\infty<\mu<+\infty$

عندئذٍ:

$$q(x_1,x_2,\dots,x_n|\mu)=\frac{1}{(2\pi)^{\frac{n+1}{2}}}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}x_i{}^2+(n+1)\mu^2-2\mu n\bar{x}\right]\right\}$$

و

$$h(x_1,x_2,\dots,x_n)=\frac{1}{(2\pi)^{\frac{n+1}{2}}}\,e^{\frac{1}{2}\sum x_i}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}[(n+1)\mu^2-2\mu n\bar{x}]}\,d\mu$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}x_i{}^2-\frac{n^2\overline{(x)}^2}{n+1}\right]\right\}$$

أو: $h(x_1,x_2,\dots,x_n)=\frac{1}{[n+1]^{\frac{1}{2}}[2\pi]^{\frac{n}{2}}}\ \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}x_i{}^2-\frac{n^2\overline{(x)}^2}{n+1}\right]\right\}$

وبالتالي:

$$k(\mu|x_1,x_2,\dots,x_n)=\frac{q(x_1,x_2,\dots,x_n|\mu)}{h(x_1,x_2,\dots,x_n)}=\left(\frac{n+1}{2\pi}\right)^2\exp\left\{-\frac{1}{2}(n+1)\left[\mu-\frac{n\bar{x}}{n+1}\right]^2\right\}$$

أي أنّ توزيع μ علماً أنّ $X_1,X_2,\dots,X_n$ معروفة هو توزيع طبيعي بمتوسط يساوي $\frac{n\bar{x}}{n+1}$ وتشتت $(n+1)^2$،

فإذا فرضنا الآن أنّ تابع الخسارة يساوي $I(\hat{\mu},\mu)=[\hat{\mu}-\mu]^2$ ، فإنّ التابع $V(\hat{\mu},x_1,x_2,\dots,x_n)$ هو:

$$V(\hat{\mu},x_1,x_2,\dots,x_n)=\int_{-\infty}^{+\infty}([\hat{\mu}-\mu]^2)\ k(\mu|x_1,x_2,\dots,x_n)d\mu$$

$$=\frac{(n+1)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}([\hat{\mu}-\mu]^2)\exp\left\{-\frac{1}{2}(n+1)\left[\mu-\frac{n\bar{x}}{n+1}\right]^2\right\}d\mu$$

$$=\widehat{\mu^2}-\frac{2\hat{\mu}\bar{x}n}{n+1}+\frac{1}{n+1}+\frac{\overline{x^2}n^3}{(n+1)^2}$$

$$\frac{dv}{d\hat{\mu}}=2\hat{\mu}-\frac{2n\bar{x}}{n+1}=0$$

ومنه تقدير بايز: $\hat{\mu}=\frac{\sum x_i}{n+1}$

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له توزيع بواسون والمعرف بدالة الكثافة:

$$f(x|\theta)=\frac{e^{-\theta}\,\theta^x}{x!}\ ;\ x=0,1,2,3,\dots\ ,\ \ \theta>0$$

وبفرض أن التوزيع السلفي لـ θ هو التوزيع الأسي والمعرف بدالة الكثافة:

$$\rho(\theta) = e^{-\theta} \;\; ; \; \theta > 0$$

وأن دالة الخسارة هي دالة الخطأ التربيعي $l(\hat{\theta},\theta) = (\hat{\theta} - \theta)^2$, والمطلوب:

1 – أوجد التوزيع اللاحق لـ θ .

2 – أوجد تقدير بايز لـ θ .

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي المعرف بدالة الكثافة:

$$f(x|\theta) = \theta x^{\theta-1} \;\; ; \; \theta < x < 1 \; , \quad \theta > 0$$

وبفرض أن التوزيع السلفي لـ θ هو توزيع غاما والمعرف بدالة الكثافة:

$$\rho(\theta) = \frac{1}{\Gamma(\beta)} \theta^{\beta-1} e^{-\theta} \;\; ; \; \theta > 0 \; , \quad \beta \text{ معلومة}$$

وأن دالة الخسارة هي دالة الخطأ التربيعي $l(\hat{\theta},\theta) = (\hat{\theta} - \theta)^2$, والمطلوب:

1 – أوجد التوزيع اللاحق لـ θ .

2 – أوجد تقدير بايز لـ θ .

مبرهنة راو – بلاك ويل The Rao – *Blackwell Theorem* :

ليكن X, Y متغيران عشوائيان وليكن μ هو التوقع الرياضي للمتغير Y وأنّ تباينه هو $Var(Y)$ ولنفرض أنّ: $E(Y \setminus x) = \varphi(x)$ عندئذٍ يكون:

1- $E[\varphi(X)] = \mu$

2- $Var(\varphi(X)) \leq Var(Y)$

البرهان:

لتكن لدينا $f(x,y), f_1(x), f_2(y), h(Y \setminus x)$ الكثافة المشتركة للمتغيرين العشوائيين X, Y, الكثافة الهامشية للمتغير X , الكثافة الهامشية للمغير Y والكثافة الشرطية للمتغير Y علماً أنّ X , على الترتيب وبالتالي يكون:

$$E(Y \setminus x) = \int_{-\infty}^{+\infty} y . h(Y \setminus x) dy \quad (1)$$

ولكن نعلم أنّ:

$$h(y \setminus x) = \frac{f(x,y)}{f_1(x)}$$

بالتعويض في (1) نجد أنّ:

$$E(Y \setminus x) = \int_{-\infty}^{+\infty} y . \frac{f(x,y)}{f_1(x)} dy = \frac{1}{f_1(x)} . \int_{-\infty}^{+\infty} y . f(x,y) dy \;\Rightarrow$$

$$\varphi(x) = \frac{1}{f_1(x)} . \int_{-\infty}^{+\infty} y . f(x,y) dy \quad \Rightarrow f_1(x) . \varphi(x) = \int_{-\infty}^{+\infty} y . f(x,y) dy$$

$$E[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) . f_1(x) dx = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} y . f(x,y) dy \right] dx$$

$$= \int_{-\infty}^{+\infty} y \left[\int_{-\infty}^{+\infty} f(x,y) dx \right] dy = \int_{-\infty}^{+\infty} y . f_2(y) dy = E(Y) = \mu$$

لنبرهن البند الثاني من النظرية السابقة:

$$Var(Y) = \sigma_Y^2 = E[Y - E(Y)]^2 = E[(Y - \mu)^2]$$

$$= E\left[\left(Y - \varphi(X) \right) + \left(\varphi(X) - \mu \right) \right]^2$$

$$= E[Y - \varphi(X)]^2 + E[\varphi(X) - \mu]^2 + 2E\left[\left(Y - \varphi(X) \right) . \left(\varphi(X) - \mu \right) \right]$$

في المجموع السابق لنثبت أنّ أخر حد يكون معدوماً وذلك حسب ما يلي:

$$E\left[\left(Y - \varphi(X) \right) . \left(\varphi(X) - \mu \right) \right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(y - \varphi(x) \right) . \left(\varphi(x) - \mu \right) . f(x,y) dx\, dy$$

$$= \int_{-\infty}^{+\infty} \left(\varphi(x) - \mu \right) . \left[\int_{-\infty}^{+\infty} \left(y - \varphi(x) \right) . h(Y \setminus x)\, dy \right] . f_1(x) dx$$

بملاحظة أنّ:

$$\int_{-\infty}^{+\infty} \left(y - \varphi(x) \right) h(Y \setminus x)\, dy = \int_{-\infty}^{+\infty} y h(Y \setminus x)\, dy - \int_{-\infty}^{+\infty} \varphi(x) h(Y \setminus x)\, dy$$

$$= \varphi(x) - \varphi(x) \int_{-\infty}^{+\infty} h(Y \setminus x)\, dy = \varphi(x) - \varphi(x) = 0$$

وبالتالي نجد أنّ:

$$Var(Y) = E[Y - \varphi(X)]^2 + E[\varphi(X) - \mu]^2$$

بملاحظة أنّ الحد الأول من الطرف الأيمن للمجموع السابق موجباً دوماً وبملاحظة أنّ:

$$E[\varphi(X) - \mu]^2 = Var\left(\varphi(X) \right)$$

نجد أنّ:

$$Var(Y) = \text{عدد موجب} + Var\left(\varphi(X) \right) \;\Rightarrow$$

$$Var(Y) \geq Var\left(\varphi(X) \right)$$

وهو المطلوب

قد تم بطريقة أو بأخرى تحديد أنّواع التقديرات وهي:

1- التقدير الغير منحاز.

2- التقدير الغير المنحاز ذو التباين الأصغري.

3- التقدير المتّسق.

4- $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ ... *

نقول أنّ $\hat{\theta}_1$ أفضل من $\hat{\theta}_2$ إذا كانا غير منحازان بالنسبة للوسيط θ والأكثر من ذلك يجب تحقق المتباينة *.

- يكون التقدير T جيد إذا حقق متباينة كرامير ـ راو أو إذا كان هذا التقدير يساوي الحد الأدنى في متباينة كرامير ـ راو .

نتيجة: أنّ المبرهنة السابقة تعني أنّنا استطعنا تحسين المقدّر Y وذلك بأخذ التوقع الرياضي الشرطي للمتغير Y عند X.

مثال: لتكن لدينا الدالة:

$$f(x,y) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}q(x,y)} \quad , \quad -\infty < x,y < \infty$$

علماً أنّ : $\sigma_1 > 0\,, \sigma_2 > 0\,, and -1 < \rho < 1$

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

حيث تعرف الثوابت: $\mu_1\,,\ \mu_2\,,\ \sigma_1^2\,, \sigma_2^2\,, and\ \rho$ وسطاء هذا التوزيع.

المطلوب:

1 – أثبت أنّ هذه الدالة هي دالة كثافة احتمالية.

2 – أثبت أنّ المتغير X له التوزيع الطبيعي $N(\mu_1, \sigma_1^2)$.

3 – أثبت أنّ المتغير Y له التوزيع الطبيعي $N(\mu_2, \sigma_2^2)$.

4 – بفرض أنّ $E(Y \setminus X = x) = \varphi(x)$, أوجد التوقع الرياضي لـ $\varphi(X)$.

5 – أثبت أنّ $Var[\varphi(x)] \le Var(Y)$.

ملاحظة: يسمى هذا التوزيع الممثل بهذه الدالة بالتوزيع الطبيعي الثنائي .

الحل:

1 – حتى تكون الدالة $f(x,y)$ هي دالة كثافة يجب أنّ تكون غير سالبة وهذا محقق لأنّها دالة أسية وأمثالها ثوابت موجبة , وأنّ يكون التكامل الثنائي لهذه الدالة بالنسبة للمتغيرين x,y مساوياً الواحد.

أي , لنثبت أنّ:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \qquad (1)$$

سوف نحاول كتابة الدالة $f(x,y)$ بشكل مبسط , وذلك لتسهيل ايجاد التكامل.

لننتم إلى مربع كامل بالنسبة للحد $\left(\frac{y-\mu_2}{\sigma_2}\right)$, أي نضيف ونطرح المقدار $\rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2$ في $q(x,y)$ فنجد :

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$
$$= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

$$= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \underbrace{\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2}_{\text{مربع كامل}}\right]$$

$$= \frac{1}{1-\rho^2}\left\{(1-\rho^2)\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left[\frac{y-\mu_2}{\sigma_2} - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right]^2\right\}$$

$$= \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{\sigma_2^2(1-\rho^2)}\left[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2$$

$$= \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{\sigma_2^2(1-\rho^2)}[y-b]^2$$

$$= \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-b}{\sigma_3}\right)^2$$

حيث : $\sigma_3 = \sigma_2\sqrt{(1-\rho^2)}$, $b = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$.

بالتعويض في العلاقة (1) نجد أنّ :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}q(x,y)}dy\right)dx$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{1}{(2\pi)\sigma_1\sigma_3\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-b}{\sigma_3}\right)^2\right]}dy\right)dx$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\left(\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}}_{N(b,\sigma_3^2)}dy\right)dx$$

وبما أنّ : $\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}dy = 1$ فإنّ :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}dx = 1$$

تسمى الدالة $f(x,y)$ المعرفة بالعلاقة السابقة بدالة التوزيع الطبيعي الثنائي.

2 – لنوجد الآن دالة الكثافة الهامشية للمتغير X , وذلك بمكاملة الدالة $f(x,y)$ بالنسبة لـ y .

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty}\frac{1}{(2\pi)\sigma_1\sigma_3\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-b}{\sigma_3}\right)^2\right]}dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}}_{N(b,\sigma_3^2)}dy = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

أي أنّ دالة الكثافة الهامشية للمتغير X تصبح بالشكل:

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \quad , \quad -\infty < x < \infty , \mu_1 \in R , \sigma_1 > 0$$

أي أنّ المتغير X له التوزيع الطبيعي $N(\mu_1, \sigma_1^2)$.

3 – لإيجاد دالة الكثافة الهامشية للمتغير Y , نعيد الخطوات السابقة وذلك بالإتمام إلى مربع كامل بالنسبة للحد $\left(\frac{x-\mu_1}{\sigma_1}\right)$, أي نضيف ونطرح المقدار $\rho^2\left(\frac{y-\mu_2}{\sigma_2}\right)^2$ في $q(x,y)$, ثم نكامل الدالة $f(x,y)$ بالنسبة لـ x فنجد :

$$f_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2} \quad , \quad -\infty < y < \infty \,, \mu_2 \in R \,, \sigma_2 > 0$$

أي أنّ المتغير Y له التوزيع الطبيعي $N(\mu_2, \sigma_2^2)$.

4 – لإيجاد التوقع الرياضي الشرطي لـ Y عند $X = x$ نوجد أولاً دالة الكثافة الشرطية $h(y \setminus x)$.

$$h(y \setminus x) = \frac{f(x,y)}{f_1(x)} = \frac{\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}}{\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}$$

وبالتالي فإنّ التوقع الرياضي الشرطي يعطى بالشكل:

$$E(Y \setminus X = x) = \int_{-\infty}^{\infty} y h(y \setminus x) = \int_{-\infty}^{\infty} (y) \underbrace{\frac{1}{\sqrt{2\pi}\sigma_3\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_3}\right)^2}}_{N(b,\sigma_3^2)} dy$$

$$= b = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) = \varphi(X)$$

لأنّ دالة الكثافة الشرطية هي كثافة احتمالية لها التوزيع الطبيعي $N(b, \sigma_3^2)$, وبالتالي فإنّ التوقع الرياضي الشرطي يساوي $b = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$.

لنوجد الآن التوقع الرياضي للدالة $\varphi(X)$ بالشكل التالي:

$$E[\varphi(X)] = E\left[\underbrace{E(Y \setminus X = x)}_{b}\right] = E(b) = E\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)\right]$$

$$= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}E(X - \mu_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(\mu_1 - \mu_1) = \mu_2$$

مما يدل على أنّ $\varphi(X)$ مقدِّر غير منحاز بالنسبة لـ μ_2 , ولدينا Y أيضاً مقدِّر غير منحاز بالنسبة لـ μ_2 , لذلك نقارن بين هذين المقدّرين , من أجل معرفة المقدِّر الأفضل وذلك بحساب تباين كل منهما , ويكون المقدِّر الأفضل هو المقدِّر الذي يكون تباينه أقل.

5 – $Var(\varphi(X)) = Var\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)\right)$

$$= 0 + \rho^2 \frac{\sigma_2^2}{{\sigma_1}^2} Var(X) = \rho^2 \frac{\sigma_2^2}{\sigma_1^2}(\sigma_1^2) = \rho^2 \sigma_2^2$$

ولدينا $Var(Y) = {\sigma_2}^2$.

نلاحظ أنّ : $Var\big(\varphi(X)\big) = \rho^2 \sigma_2^2 \le \sigma_2^2 = Var(Y)$

أي أنّ المقدِّر $\varphi(X)$ غير المنحاز بالنسبة لـ μ_2 أفضل من المقدِّر Y غير المنحاز بالنسبة لـ μ_2 . وبهذا نكون قد استطعنا تحسين المقدِّر Y *بأخذ التوقع الشرطي* $E(Y \setminus X = x)$ والذي بدوره يكون تابعاً للمتغير X نسميه $\varphi(X)$.

مثال: ليكن X, Y متغيرين عشوائيين, دالة كثافتهما المشتركة :

$$f_{X,Y}(x,y) = \frac{2}{\theta^2} e^{-\frac{x+y}{\theta}} \qquad ;0 < x < y < \infty \quad ,\theta > 0$$

والمطلوب:

1. أوجد دالة الكثافة الهامشية لكل من Y و X .
2. أوجد $E(Y)$ و $Var(Y)$.
3. بفرض أنّ : $E(Y/x) = \varphi(x)$ أثبت أنّ $Var(\varphi(x)) \le Var(Y)$.

الحل:

$$f_X(x) = \int_{y=x}^{y=\infty} f_{X,Y}(x,y)dy = \int_{y=x}^{y=\infty} \frac{2}{\theta^2} e^{-\frac{x+y}{\theta}} dy$$
$$= \frac{2}{\theta} e^{-\frac{x}{\theta}} \int_{y=x}^{y=\infty} \frac{1}{\theta} e^{-\frac{y}{\theta}} dy = \frac{2}{\theta} e^{-\frac{x}{\theta}} \left[-e^{-\frac{y}{\theta}}\right]_{y=x}^{y=\infty}$$
$$= \frac{2}{\theta} e^{-\frac{x}{\theta}} \left[e^{-\frac{x}{\theta}}\right] = \frac{2}{\theta} e^{-\frac{2x}{\theta}} \qquad ;\theta > 0, x > 0$$

أي أنّ X أسي وسيطه $\lambda = \frac{2}{\theta}$.

$$f_y(y) = \int_{x=0}^{x=y} f_{X,Y}(x,y)dx = \int_{x=0}^{x=y} \frac{2}{\theta^2} e^{-\frac{x+y}{\theta}} dx$$
$$= \frac{2}{\theta} e^{-\frac{y}{\theta}} \int_{x=0}^{x=y} \frac{1}{\theta} e^{-\frac{x}{\theta}} dy = \frac{2}{\theta} e^{-\frac{y}{\theta}} \left[-e^{-\frac{x}{\theta}}\right]_{x=0}^{x=y}$$
$$= \frac{2}{\theta} e^{-\frac{y}{\theta}} \left[1 - e^{-\frac{y}{\theta}}\right] = \frac{2}{\theta} e^{-\frac{y}{\theta}} - \frac{2}{\theta} e^{-\frac{2y}{\theta}} \qquad ;\theta > 0, y > 0$$

لنوجد التوقع الرياضي:

$$E(Y) = \int_0^\infty y f_y(y)dy = \int_0^\infty y\left[\frac{2}{\theta} e^{-\frac{y}{\theta}} - \frac{2}{\theta} e^{-\frac{2y}{\theta}}\right] dy = \frac{2}{\theta} \underbrace{\int_0^\infty y\, e^{-\frac{y}{\theta}} dy}_{I_1} - \frac{2}{\theta} \underbrace{\int_0^\infty y\, e^{-\frac{2y}{\theta}} dy}_{I_2}$$

$$I_1 = \int_0^\infty y\, e^{-\frac{y}{\theta}} dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{1}{\theta}y = \mathfrak{z} \Rightarrow dy = \theta d\mathfrak{z}$$

وبالتالي فإنّ :

$$I_1 = \theta^2 \int_0^\infty \mathfrak{z}\, e^{-\mathfrak{z}} d\mathfrak{z} = \theta^2 \int_0^\infty \mathfrak{z}^{2-1}\, e^{-\mathfrak{z}} d\mathfrak{z} = \theta^2 \Gamma(2) = \theta^2$$

حساب التكامل الثاني:

$$I_2 = \int_0^\infty y\, e^{-\frac{2y}{\theta}} dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{2}{\theta}y = t \Rightarrow dy = \frac{\theta}{2}dt$$

وبالتالي فإنّ :

$$I_2 = \frac{\theta^2}{4}\int_0^\infty t\, e^{-t}dt = \frac{\theta^2}{4}\int_0^\infty t^{2-1}\, e^{-t}dt = \frac{\theta^2}{4}\Gamma(2) = \frac{\theta^2}{4}$$

$$E(y) = 2.\frac{1}{\theta}.\theta^2 - 2.\frac{1}{\theta}.\theta^4 = 2\theta - \frac{1}{2}\theta = \frac{3}{2}\theta \quad Var(y) = \underbrace{Ey^2}_{?} - [E(y)]^2$$

$$E(y^2) = \int_0^\infty y^2 f_y(y)dy = \int_0^\infty y^2 \left[\frac{2}{\theta}\, e^{-\frac{y}{\theta}} - \frac{2}{\theta}\, e^{-\frac{2y}{\theta}}\right] dy$$

$$= \frac{2}{\theta}\underbrace{\int_0^\infty y^2\, e^{-\frac{y}{\theta}}dy}_{I_1} - \frac{2}{\theta}\underbrace{\int_0^\infty y^2\, e^{-\frac{2y}{\theta}}dy}_{I_2}$$

$$I_1 = \int_0^\infty y^2\, e^{-\frac{y}{\theta}}dy$$

$$\frac{1}{\theta}y = \mathfrak{z} \Rightarrow dy = \theta d\mathfrak{z} \Rightarrow I_1 = \theta^3 \int_0^\infty \mathfrak{z}^2\, e^{-\mathfrak{z}} d\mathfrak{z} = \theta^3 \int_0^\infty \mathfrak{z}^{3-1}\, e^{-\mathfrak{z}} d\mathfrak{z} = \theta^3 \Gamma(3) = 2\theta^3$$

$$I_2 = \int_0^\infty y^2\, e^{-\frac{2y}{\theta}}dy$$

$$\frac{2}{\theta}y = t \Rightarrow dy = \frac{\theta}{2}dt \Rightarrow I_2 = \frac{\theta^3}{8}\int_0^\infty t^2\, e^{-t}dt = \frac{\theta^3}{8}\int_0^\infty t^{3-1}\, e^{-t}dt = \frac{\theta^3}{8}\Gamma(3) = \frac{\theta^3}{4}$$

$$E(y^2) = \frac{2}{\theta}\, 2\theta^3 - \frac{2}{\theta}\,\frac{\theta^3}{4} = 4\theta^2 - \frac{1}{2}\theta^2 = \frac{7}{2}\theta^2$$

$$Var(y) = Ey^2 - [E(y)]^2 = \frac{7}{2}\theta^2 - \left(\frac{3}{2}\theta^2\right)^2 = \frac{5}{4}\theta^2$$

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{2}{\theta^2}\, e^{-\frac{x+y}{\theta}}}{\frac{2}{\theta}\, e^{-\frac{2x}{\theta}}} = \frac{1}{\theta}\, e^{-\frac{1}{\theta}(y-x)}\ ; \begin{cases} 0 < x < y < \infty \\ \quad 0 < x \end{cases}$$

$$E(y/x) = \int_{y=x}^{y=\infty} y\, f_{Y/X}(y/x)dy = \int_{y=x}^{y=\infty} y \left[\frac{1}{\theta}\, e^{-\frac{1}{\theta}(y-x)}\right] dy$$

$$= \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{y=x}^{y=\infty} y\, e^{-\frac{1}{\theta}y}dy = \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} \theta^2\, \mathfrak{z}\, e^{-\mathfrak{z}} d\mathfrak{z}$$

$$= \theta\, e^{\frac{x}{\theta}} \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} \mathfrak{z}\, e^{-\mathfrak{z}} d\mathfrak{z}$$

نكامل بالتجزئة لذلك نفرض :

$$\left.\begin{aligned} u &= \mathfrak{z} \Rightarrow du = d\mathfrak{z} \\ dv &= e^{-\mathfrak{z}} d\mathfrak{z} \Rightarrow v = -\, e^{-\mathfrak{z}} \end{aligned}\right\} \Rightarrow E(y/x) = \theta\, e^{\frac{x}{\theta}} \left[-\mathfrak{z}\, e^{-\mathfrak{z}} \Big|_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} + \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} e^{-\mathfrak{z}} d\mathfrak{z} \right]$$

$$= \theta\, e^{\frac{x}{\theta}}\left[\frac{x}{\theta}\, e^{-\frac{x}{\theta}} + e^{-\frac{x}{\theta}}\right] = x + \theta = \varphi(x)$$

$$Var(\varphi(x)) = \underbrace{E[\varphi(x)]^2}_{?} - \underbrace{(E[\varphi(x)])^2}_{?}$$

$$E[\varphi(x)] = E[x+\theta] = E[x] + \theta = \frac{\theta}{2} + \theta = \frac{3}{2}\theta$$

$$E[\varphi(x)]^2 = E[x+\theta]^2 = E[x^2 + 2\theta x + \theta^2]$$

$$= E[x^2] + 2\theta E[x] + \theta^2$$

$$= E[x^2] + 2\theta\left[\frac{\theta}{2}\right] + \theta^2$$

$$= \frac{\theta^2}{2} + \theta^2 + \theta^2 = \frac{5}{2}\theta^2$$

حيث لدينا X أسي وسيطه $\lambda = \frac{2}{\theta}$.

$$\Rightarrow Var(X) = E[X^2] - [E(X)]^2 \Rightarrow$$

$$E[X^2] = Var(X) + [E(X)]^2$$

$$= \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2 = \frac{\theta^2}{4} + \frac{\theta^2}{4} = \frac{\theta^2}{2} \Rightarrow$$

$$Var(\varphi(X)) = E[\varphi(X)]^2 - [E[\varphi(X)]]^2 = \frac{5}{2}\theta^2 - \left(\frac{3}{2}\theta\right)^2 = \frac{5}{2}\theta^2 - \frac{9}{4}\theta^2 = \frac{1}{4}\theta^2$$

نلاحظ أنّ :

$$Var(Y) = \frac{5}{4}\theta^2 > \frac{1}{4}\theta^2 = Var(\varphi(X))$$

أي أنّ التقدير غير المنحاز $\varphi(X)$ أفضل من التقدير غير المنحاز Y

مبرهنة هامة للتمارين: الشرط اللازم والكافي لكي تسمح الدالة $L(\underline{x}\,,\theta)$ بتقدير يُدرك تباينه الحد الأدنى في متباينة كرامير – راو (حد المعلومات) من أجل تابع في θ نختاره بشكل معين هو أنّ يكون من الشكل:

$$L(\underline{x}\,,\theta) = \mathrm{L} = \mathrm{L}_1 . e^{t\theta_1 + \theta_2}$$

حيث t_1 و t تابعان في $x_1, x_2, \dots x_n$ فقط و θ_1 و θ_2 تابعان في θ فقط.

تعريف:

الإحصاء المرتب: لتكن لدينا $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي له توزيع مستمر ومعرف بدالة الكثافة الاحتمالية $f(x,\theta)$, حيث $a < x < b$.

لنفرض أنّ المتغير Y_1 هو أصغر متغيرات العينة العشوائية $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$, والمتغير Y_2 هو أصغر متغيرات العينة العشوائية $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$ بعد استبعاد الأصغر وهكذا ... , وأنّ Y_n هو أكبر متغيرات العينة $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$ بحيث نحصل على الإحصاء $Y_1 < Y_2 < \cdots < Y_n$. عندئذ نسمي الإحصاء Y_i , حيث $i = 1,2, \dots, n$ بالإحصاء المرتب الموافق للعينة العشوائية $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$. وتكون:

1 – دالة الكثافة المشتركة للمتغيرات العشوائية $Y_1, Y_2, \dots, Y_n$ معرفة بالعلاقة:

$$g(y_1,y_2,\dots,y_n)=(n!)f(y_1)f(y_2)\dots f(y_n)\quad ;\quad a<y_1<y_2<\dots<y_n<b$$

2 – دالة الكثافة الهامشية للمتغير الأكبر Y_n , $g_n(y_n)$ تعطى بالعلاقة:

$$g_n(y_n)=n[F(y_n)]^{n-1}f(y_n)\quad ;\quad a<y_n<b$$

3 – دالة الكثافة الهامشية للمتغير الأصغر Y_1 , $g_1(y_1)$ تعطى بالعلاقة:

$$g_1(y_1)=n[1-F(y_1)]^{n-1}f(y_1)\quad ;\quad a<y_1<b$$

4 – دالة الكثافة الهامشية للمتغير Y_k , $g_k(y_k)$ تعطى بالعلاقة:

$$g_k(y_k)=\frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k)\quad ;\quad a<y_k<b$$

5 – دالة الكثافة الهامشية للمتغيرين العشوائيين Y_i,Y_j , $g_{ij}(y_i,y_j)$ تعطى بالعلاقة:

$$g_{ij}(y_i,y_j)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}$$
$$\times[F(y_i)]^{i-1}[F(y_j)-F(y_i)]^{j-i-1}[1-F(y_j)]^{n-j}f(y_i)f(y_j)\quad ;\quad a<y_i<y_j<b$$

مثال: لتكن $X_1,X_2,\dots,X_7$ عينة عشوائية حجمها $n=7$ والمأخوذة من مجتمع إحصائي له التوزيع المنتظم على الفترة $[0,\theta]$ والمعرف بدالة الكثافة التالية:

$$f(x,\theta)=\frac{1}{\theta}\quad ;\quad 0<x<\theta\,,0<\theta<\infty$$

وليكن $Y_1<Y_2<Y_3<Y_4<Y_5<Y_6<Y_7$ الإحصاء المرتب الموافق لهذه العينة, والمطلوب:

1 – أوجد دالة الكثافة للمتغير العشوائي Y_4 ثم للمتغير العشوائي Y_7 .

2 – أثبت أنّ Y_7 هو إحصاء كاف بالنسبة للوسيط θ .

3 – أثبت أنّ المقدِّر $T=\left(\frac{n+1}{4}\right)Y_4$ هو مقدِّر غير منحاز بالنسبة للوسيط θ واحسب تباينه .

4 – أثبت أنّ المقدِّر $T=\left(\frac{n+1}{4}\right)Y_4$ هو مقدِّر متسق.

5 – أوجد دالة الكثافة الشرطية لـ Y_4 عند $Y_7=y_7$, ثم أوجد التوقع الرياضي الشرطي لـ Y_4 عند $Y_7=y_7$.

6 – بفرض أنّ: $E(2Y_4|Y_7)=\varphi(Y_7)$, قارن بين المقدِّرين $\varphi(Y_7)$ و T .

الحل:

لدينا حسب تعريف الإحصاء المرتب أن دالة الكثافة الهامشية للمتغير Y_k , $g_k(y_k)$ تعطى بالعلاقة:

$$g_k(y_k)=\frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k)\quad ;\quad a<y_k<b$$

أي أن:

$$g_4(y_4)=\frac{7!}{(3)!(3)!}\left[\frac{y_4}{\theta}\right]^3\left[1-\frac{y_4}{\theta}\right]^3\frac{1}{\theta}$$
$$=(140)\left(\frac{1}{\theta}\right)^4(y_4^3)\left(1-\frac{y_4}{\theta}\right)^3\quad ;\quad 0<y_4<\theta$$

كما ولدينا أن دالة الكثافة الهامشية للمتغير الأكبر Y_n , $g_n(y_n)$ تعطى بالعلاقة:

$$g_n(y_n) = n[F(y_n)]^{n-1} f(y_n) \quad ; \quad a < y_n < b$$

أي أن:

$$g_7(y_7) = 7[F(y_7)]^6 f(y_7) = 7\left(\frac{y_7}{\theta}\right)^6 \left(\frac{1}{\theta}\right) = \frac{7}{\theta^7} y_7^6 \quad ; \quad 0 < y_7 < \theta$$

2 – لنثبت أنّ Y_7 هو إحصاء كاف بالنسبة للوسيط θ .

$$L(\underline{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{7} \left(\frac{1}{\theta}\right) = \frac{1}{\theta^7} = \left(\frac{7y_7^6}{\theta^7}\right)\left(\frac{1}{7y_7^6}\right)$$

$$= \left(\frac{7y_7^6}{\theta^7}\right)\left(\frac{1}{7}\frac{1}{(max\{X_1, X_2, \dots, X_7\})^2}\right)$$

$$= \underbrace{g_7(y_7)}_{Y_7 \text{ كثافة}} \quad \underbrace{H(x_1, x_2, \dots, x_7)}_{\text{دالة تتعلق بالمتغيرات } X_1, X_2, \dots, X_7 \text{ فقط}}$$

وبالتالي فإن الإحصاء Y_7 هو إحصاء كاف بالنسبة للوسيط θ .

3 – لنوجد التوقع الرياضي والتباين للمقدِّر T .

$$E(Y_4) = \int_0^\theta y_4 g_4(y_4) dy_4 = \int_0^\theta y_4 (140) \left(\frac{1}{\theta}\right)^4 (y_4^3) \left(1 - \frac{y_4}{\theta}\right)^3 dy_4$$

$$= (140)\left(\frac{1}{\theta}\right)^4 \int_0^\theta y_4^4 \left(1 - 3\frac{y_4}{\theta} + 3\frac{y_4^2}{\theta^2} - \frac{y_4^3}{\theta^3}\right) dy_4$$

$$= (140)\left(\frac{1}{\theta}\right)^4 \int_0^\theta \left(y_4^4 - 3\frac{y_4^5}{\theta} + 3\frac{y_4^6}{\theta^2} - \frac{y_4^7}{\theta^3}\right) dy_4$$

$$= (140)\left(\frac{1}{\theta}\right)^4 \left(\frac{y_4^5}{5} - \frac{3y_4^6}{6\theta} + \frac{3y_4^7}{7\theta^2} - \frac{y_4^8}{8\theta^3}\right)\Big|_0^\theta$$

$$= (140)\left(\frac{1}{\theta}\right)^4 \left(\frac{\theta^5}{5} - \frac{3\theta^6}{6\theta} + \frac{3\theta^7}{7\theta^2} - \frac{\theta^8}{8\theta^3}\right)$$

$$= (140)\left(\frac{1}{\theta}\right)^4 \left(\frac{\theta^5}{5} - \frac{\theta^5}{2} + \frac{3\theta^5}{7} - \frac{\theta^5}{8}\right)$$

$$= (140)\left(\frac{1}{5} - \frac{1}{2} + \frac{3}{7} - \frac{1}{8}\right)\theta = \frac{\theta}{2}$$

وبالتالي فإن:

$$E(T) = E\left[\left(\frac{n+1}{4}\right) Y_4\right] = E\left[\left(\frac{7+1}{4}\right) Y_4\right] = E(2Y_4) = \theta$$

أي أن T مقدِّر غير منحاز بالنسبة للوسيط θ .

لإيجاد تباين T نوجد التوقع الرياضي لـ T^2 , ثم نعوض في علاقة التباين:

$$E(T)^2 = E\left[\left(\frac{n+1}{4}\right) Y_4\right]^2 = E\left[\left(\frac{7+1}{4}\right) Y_4\right]^2 = E(2Y_4)^2$$

$$= \int_0^\theta 4y_4^2 g_4(y_4) dy_4 = \int_0^\theta 4y_4^2 (140) \left(\frac{1}{\theta}\right)^4 (y_4^3) \left(1 - \frac{y_4}{\theta}\right)^3 dy_4$$

$$= 560 \left(\frac{1}{\theta}\right)^4 \int_0^\theta y_4^5 \left(1 - \frac{y_4}{\theta}\right)^3 dy_4$$

$$= 560\left(\frac{1}{\theta}\right)^4 \int_0^\theta y_4^5\left(1-3\frac{y_4}{\theta}+3\frac{y_4^2}{\theta^2}-\frac{y_4^3}{\theta^3}\right) dy_4$$
$$= 560\left(\frac{1}{\theta}\right)^4 \int_0^\theta \left(y_4^5-3\frac{y_4^6}{\theta}+3\frac{y_4^7}{\theta^2}-\frac{y_4^8}{\theta^3}\right) dy_4$$
$$= 560\left(\frac{1}{\theta}\right)^4 \left(\frac{y_4^6}{6}-\frac{3y_4^7}{7\theta}+\frac{3y_4^8}{8\theta^2}-\frac{y_4^9}{9\theta^3}\right)\Big|_0^\theta$$
$$= 560\left(\frac{1}{\theta}\right)^4 \left(\frac{\theta^6}{6}-\frac{3\theta^7}{7\theta}+\frac{3\theta^8}{8\theta^2}-\frac{\theta^9}{9\theta^3}\right)$$
$$= 560\left(\frac{1}{\theta}\right)^4 \left(\frac{\theta^6}{6}-\frac{3\theta^6}{7}+\frac{3\theta^6}{8}-\frac{\theta^6}{9}\right)$$
$$= 560\left(\frac{1}{6}-\frac{3}{7}+\frac{3}{8}-\frac{1}{9}\right)\theta^2 = \left(\frac{70}{63}\right)\theta^2$$

ومنه نجد:

$$Var(T) = E(T)^2 - [E(T)]^2 = \left(\frac{70}{63}\right)\theta^2 - \theta^2 = \frac{1}{9}\theta^2$$

4 – نلاحظ أن:

$$Var(T) = \frac{1}{9}\theta^2 = \frac{\theta^2}{n+2} \xrightarrow[n\to\infty]{} 0$$

وبما أن المقدِّر T هو مقدِّر غير منحاز بالنسبة للوسيط θ , وأن تباينه يسعى للصفر عندما تسعى n إلى اللانهاية , فإن المقدِّر T يكون متسقاً .

5 –

مثال: لتكن $X_1, X_2, \dots, X_5$ عينة عشوائية حجمها $n = 5$ والمأخوذة من مجتمع إحصائي له التوزيع المنتظم على الفترة $[0, \theta]$ والمعرف بدالة الكثافة التالية:

$$f(x,\theta) = \frac{1}{\theta} \quad ; \quad 0 < x < \theta \,, 0 < \theta < \infty$$

وليكن $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ الإحصاء المرتب الموافق لهذه العينة, والمطلوب:

1 – أوجد دالة الكثافة للمتغير العشوائي Y_3.

2 – أثبت أنّ المقدِّر $2Y_3$ هو مقدِّر غير منحاز بالنسبة للوسيط θ .

3 – أثبت أنّ Y_5 هو إحصاء كاف بالنسبة للوسيط θ .

4 – أوجد التوقع الرياضي الشرطي لـ $2Y_3$ عند $Y_5 = y_5$.

5 – بفرض أنّ: $E(2Y_3|Y_5) = \varphi(Y_5)$.

الحل:

نلاحظ أنّ : $f(x,\theta) = \frac{1}{\theta}$

$$g_3(y_3,\theta) = \int_0^{y_3}\int_{y_2}^{y_3}\int_{y_3}^{\theta}\int_{y_4}^{\theta} g(y_1,y_2,\dots,y_5,\theta)dy_1dy_2dy_4dy_5$$

$$= \frac{y_3^3}{2}\left[\frac{(\theta-y_4)^2}{2}\right]\Bigg|_{y_3}^{\theta} \cdot \frac{5!}{\theta^5}$$

$$= \frac{5!}{\theta^5} \cdot \frac{y_3^2}{2} \cdot \frac{(\theta-y_3)^2}{2} = \frac{30}{\theta^5}(\theta^2 y_3^2 - 2\theta y_3^3 + y_3^4)$$

$$g_5(y_1, \dots, y_5) = f(y_1,\theta) \dots \dots f(y_5,\theta). 5! = \frac{5!}{\theta^5}$$

كي يكون $2Y_3$ مقدِّر غير منحاز يجب أنّ يكون $E(2Y_3) = \theta$

$$\mathrm{E}(2\mathrm{Y}_3) = \frac{30}{\theta^5}\int_0^{\theta} 2\mathrm{y}_3(\theta^2 y_3^2 - 2\theta y_3^3 + y_3^4)dy_3 = \frac{30}{\theta^5}.\theta^6\left(\frac{2}{4} - \frac{4}{5} + \frac{2}{6}\right) = \theta$$

إذن $2\mathrm{Y}_3$ فعلاً هو مقدار غير متحيز لـ θ .

$$\mathrm{E}(2\mathrm{Y}_3)^2 = \frac{30}{\theta^5}\int_0^{\theta} 4\mathrm{Y}_3^2(\theta^2 y_3^2 - 2\theta y_3^3 + y_3^4)dY_3 = \frac{30}{\theta^5}.\theta^7\left(\frac{4}{5} - \frac{8}{6} + \frac{4}{7}\right) = \frac{8}{7}\theta^2$$

$$Var(2\mathrm{Y}_3) = \frac{8}{7}\theta^2 - \theta^2 = \frac{\theta^2}{7}$$

ونريد أنّ نحسن هذا المقدِّر ولذلك نأخذ Y_5 إحصاء وهو ليس تابع لـ $2\mathrm{Y}_3$ لوحدها.

أيضاً $2\mathrm{Y}_3$ نلاحظ أنّها غير تابعة لـ Y_5 .

$$\mathrm{g}(y_1, y_2, \dots, y_5) = \frac{5!}{\theta^5} = \frac{\max\{x_i\}}{\theta^5}. 5! \times \frac{1}{\max\{x_i\}} = \mathrm{k}_1(\theta, y_1).\mathrm{k}_2(\underline{x})$$

إذاً Y_5 إحصاء كافٍ لـ θ .

$$E(2\mathrm{Y}_3|Y_5) = \varphi(Y_5)$$

أنّ $\varphi(Y_5)$ مقدِّر غير منحاز لـ θ , ولذلك نحسب $Var\ (\varphi(Y_5))$.

لإيجاد التوقع الرياضي لـ $\varphi(Y_5)$ نوجد الكثافة $h(Y_3|Y_5) = \frac{g_{3,5}(Y_3,Y_5)}{g_5(Y_5)}$, لحساب تابع الكثافة لـ Y_5 نحسب الشرطية.

$$G\ Y_5(y_5) = P(Y_5 < y_5) = P\{[x_1 < y_5] \cap \dots \cap [x_5 < y_5]\}$$

$$= \{P[x < y_5]\}^5 = \left\{\frac{y_5}{\theta}\right\}^5$$

إذن : $G\ Y_5(y_5) = \frac{5}{\theta^5}y_5^4$

كذلك : $g(Y_3|Y_5) = \iiint_B\ g(y_1, y_2, \dots, y_n)dy_1 dy_2 dy_4$

حيث B هي الرقعة المناسبة في R^3 .

$$g(Y_3|Y_5) = \frac{5!}{\theta^5}\int_0^{y_3} dy_2 \int_0^{y_3} dy_2 \int_{y_3}^{y_5} dy_4 = \frac{5!}{\theta^5}\frac{y_3^2}{2}(y_5 - y_3)$$

$$h(Y_3|Y_5) = \frac{5!}{\theta^5}\frac{\frac{y_3^2}{2}(y_5-y_3)}{\frac{5}{\theta^5}y_3^4} = \frac{4!}{2y_5^4}y_3^2(y_5 - y_3)$$

الآن لدي الأمل المشروط :

$$\varphi(Y_5) = E(Y_3|Y_5) = \int_0^{y_5} h(Y_3|Y_5)dy_3$$

$$= \frac{4!}{y_5^4}\int_0^{y_5} y_3 \ (y_5^3 - y_3^4)dy_3 = \frac{4!}{y_5^4} y_5^5 \left(\frac{1}{4} - \frac{1}{5}\right) = \frac{6}{5} y_5$$

إذن : $E\left(\varphi(y_5)\right) = E\left(\frac{6}{5}y_5\right) = \theta$

ولنحسب الأمل الرياضي لمربعه وذلك لحساب تباينه:

$$E[\varphi(y_5)]^2 = \int_0^{\theta}\left(\frac{6}{5}y_5\right)^2 \frac{5}{\theta^5} y_5^4 dy_5 = \frac{36}{35}\theta^2$$

$$Var[\varphi(y_5)] \quad = \frac{36}{35}\theta^2 - \theta^2 = \frac{\theta^2}{35}$$

وهكذا نرى أنّنا حسنا هذا المقدار بشكل كبير : أي أنّ:

$$Var[\varphi(y_5)] = \frac{\theta^2}{35} < \frac{\theta^2}{7} = Var\ (2y_3)$$

وهكذا نرى أنّنا قد حسّنا التقدير باستخدام الإحصاء الكافي وذلك بالشكل $\varphi(y_5) = E(2Y_3|Y_5)$.

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي , دالة كثافته الاحتمالية معرفة بالشكل التالي:

$$f(x,\theta) = \theta x^{\theta-1} = e^{(\theta Ln\, x - Ln\, x + Ln\, \theta)} \quad ; \ 0 < x < 1 , \theta > 0$$

والمطلوب:

1 – بفرض أنّ $Y_i = -Ln(X_i)$, أوجد دالة الكثافة الاحتمالية للمتغير العشوائي $Y_i = -Ln(X_i)$.

2 – أوجد دالة الكثافة الاحتمالية للمتغير العشوائي $Z = -\sum_{i=1}^{n} Ln(X_i)$.

3 – عين الثابتة a كي يكون المقدّر $U = \frac{a}{Z}$ هو مقدّر غير منحاز بالنسبة للوسيط θ .

4 – أوجد الحد الأدنى للدالة $f(x,\theta)$.

5 – قارن بين تباين المقدّر غير المنحاز U والحد الأدنى .

الحل:

1 – لنفرض أنّ التحويل:

$$y_i = -Ln(x_i) \Rightarrow -y_i = Ln(x_i) \Rightarrow x_i = e^{-y_i}$$

هو تحويل مطابق أي تحويل واحد لواحد من الفضاء $\{x_i : 0 < x_i < 1\}$ إلى الفضاء $\{y_i : 0 < y_i < \infty\}$.

$$\frac{dx_i}{dy_i} = -e^{-y_i} \Rightarrow |J| = \left|\frac{dx_i}{dy_i}\right| = \left|-e^{-y_i}\right| = e^{-y_i}$$

وبالتالي فإنّ دالة الكثافة الاحتمالية للمتغير العشوائي Y_i تصبح بالشكل:

$$g(y_i) = f(x_i)\left|\frac{dx_i}{dy_i}\right|_{x_i = e^{-y_i}} = \theta(e^{-y_i})^{\theta-1}e^{-y_i}$$

$$= \theta e^{-\theta y_i} e^{y_i} e^{-y_i} = \theta e^{-\theta y_i}$$

$$= \frac{\theta^1}{\Gamma(1)} {y_i}^{1-1} e^{-\theta y_i} \quad , \quad y_i > 0 \quad , \quad \alpha = 1 , \beta = \frac{1}{\theta}$$

وهذه الدالة هي دالة كثافة لتوزيع غاما وسيطاه $\alpha = 1 , \beta = \frac{1}{\theta}$.

2 – بما أنّ $X_1, X_2, \dots, X_n$ هي متغيرات عشوائية مستقلة, فإنّ $Y_1, Y_2, \dots, Y_n$ هي متغيرات عشوائية مستقلة , وأنّها تخضع لتوزيع غاما وسيطاه $\alpha = 1 , \beta = \frac{1}{\theta}$, كما وأنّ المتغير $Z = -\sum_{i=1}^{n} Ln(X_i) = \sum_{i=1}^{n} Y_i$ هو متغير عشوائي له توزيع غاما وسيطاه $\alpha = n , \beta = \frac{1}{\theta}$.

وبالتالي فإنّ التوقع الرياضي والتباين للمتغير العشوائي Z تصبحان بالشكل التالي:

$$E(Z) = \alpha\beta = \frac{n}{\theta} \quad , \quad Var(Z) = \alpha\beta^2 = \frac{n}{\theta^2}$$

3 – لإيجاد قيمة a كي يكون المقدِّر $U = \frac{a}{Z}$ هو مقدِّر غير منحاز بالنسبة للوسيط θ , نوجد التوقع الرياضي للمتغير $U = \frac{a}{Z}$ ونجعله مساوياً الوسيط θ , فنجد:

$$E(U) = E\left(\frac{a}{Z}\right) = \int_0^\infty \left(\frac{a}{z}\right) \frac{\theta^n}{\Gamma(n)} z^{n-1} e^{-\theta z} dz = a \int_0^\infty \frac{\theta^n}{\Gamma(n)} z^{n-2} e^{-\theta z} dz$$

نجري تحويل في المتغير , فنفرض أنّ :

$$\theta z = y \Longrightarrow z = \frac{1}{\theta} y \Longrightarrow dz = \frac{dy}{\theta}$$

فنجد:

$$E(U) = a \frac{\theta^n}{\Gamma(n)} \int_0^\infty \left(\frac{y}{\theta}\right)^{n-2} e^{-y} \frac{dy}{\theta} = a \frac{\theta^n}{\theta^{n-1}\Gamma(n)} \int_0^\infty y^{n-2} e^{-y} dy$$

$$= a \frac{\theta \Gamma(n-1)}{\Gamma(n)} = \frac{a\theta}{n-1} \equiv \theta \Longrightarrow a = n - 1$$

4 – لنوجد الحد الأدنى في متباينة كرامير – راو للدالة $f(x, \theta)$ بالشكل التالي:

لدينا دالة الكثافة للتوزيع الطبيعي معرفة بالشكل:

$$f(x, \theta) = \theta x^{\theta - 1}$$

وبأخذ اللوغاريتم الطبيعي للطرفين نجد:

$$Lnf(x, \theta) = Ln(\theta) + (\theta - 1) Ln(x)$$

وباشتقاق الطرفين جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial Lnf(x,\theta)}{\partial \theta} = \frac{1}{\theta} + Ln(x)$$

وبالاشتقاق مرة أخرى جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial^2 Lnf(x,\theta)}{\partial \theta^2} = -\frac{1}{\theta^2}$$

لنوجد الآن معلومات فيشر من العلاقة:

$$I = E\left(-\frac{\partial^2 Lnf(x,\theta)}{\partial\theta^2}\right) = E\left(\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$$

وبالتالي فإنّ الحد الأدنى للتباين كما تحدده متباينة كرامير − راو يساوي $\frac{\theta^2}{n}$.

5 − لإيجاد التباين نوجد التوقع الرياضي لـ U^2 .

$$E(U^2) = \int_0^\infty \left(\frac{n-1}{z}\right)^2 \frac{\theta^n}{\Gamma(n)} z^{n-1} e^{-\theta z} dz = (n-1)^2 \frac{\theta^n}{\Gamma(n)} \int_0^\infty z^{n-3} e^{-\theta z} dz$$

$$= (n-1)^2 \frac{\theta^n}{\Gamma(n)} \int_0^\infty \left(\frac{y}{\theta}\right)^{n-3} e^{-y} \frac{dy}{\theta} = (n-1)^2 \frac{\theta^n}{\theta^{n-2}\Gamma(n)} \int_0^\infty y^{n-3} e^{-y} dy$$

$$= (n-1)^2 \frac{\theta^n \Gamma(n-2)}{\theta^{n-2}\Gamma(n)} = \frac{\theta^2\,(n-1)}{n-2}$$

بالتعويض نجد:

$$Var(U) = E(U^2) - [E(U)]^2 = \frac{\theta^2\,(n-1)}{n-2} - \theta^2 = \theta^2\left(\frac{n-1}{n-2} - \frac{n-2}{n-2}\right) = \frac{\theta^2}{n-2}$$

لنحسب نسبة التباين إلى الحد الأدنى فنجد:

$$\frac{Var(U)}{\text{الحد الأدنى}} = \frac{\frac{\theta^2}{n-2}}{\frac{\theta^2}{n}} = \frac{n}{n-2} \xrightarrow[n\to\infty]{} 1$$

الإحصاء التام: لتكن لدينا أسرة من التوزيعات الاحتمالية $f(x,\theta)$ بمتغير واحد أو أكثر وليكن $h(x)$ أي إحصاء لا يحوي θ فإذا كان الشرط:

$$E[h(x)] = \int h(x)\,.\,f(x,\theta)dx \equiv 0$$

من أجل جميع قيم θ يؤدي إلى أنّ: $h(x) \equiv 0$, عندها قلنا أنّ الأسرة $f(x,\theta)$ هي أسرة تامة.

وإذا كان الشرط الأول يؤدي إلى تحقق الشرط الثاني من أجل التوابع $h(x)$ **المحدودة فقط** قلنا أنّ الأسرة $f(x,\theta)$ تامة محدودة وقياساً على ذلك:

نقول أنّ الإحصاء T هو إحصاء تام إذا كان توزيعه الاحتمالي منتمياً إلى أسرة توزيعات تامة وبشكل خاص إذا كان T إحصاءً كافياً فنقول أنّه إحصاء تام إذا كان تحقق الشرط:

$E[g(T)] = 0$ من أجل جميع قيم θ يؤدي إلى أنّ: $g(T) \equiv 0$

حيث $g(T)$ هو أي تابع في T.

وبعبارة أخرى نقول عن T أنّه إحصاء تام إذا لم يكن هناك أي تقدير للصفر يمكن التعبير عنه على شكل تابع في T.

مثال: نعلم أنّه في التوزيع الحداني أنّ :

$T=\sum_{i=1}^{n} x_i$ هو إحصاء كافٍ من أجل p فهل هو إحصاء تام ؟

نعلم أنّ دالة الكثافة للمتغير T هي:

$$f(t,p)=\binom{n}{t}.p^t.(1-p)^{n-t} \quad ; \quad t=0,1,2,\dots n$$

ومنه:

$$f(t,p)=\binom{n}{t}.\left(\frac{p}{1-p}\right)^t.(1-p)^n=\binom{n}{t}.\rho^t.(1-p)^n \quad ;\rho=\frac{p}{1-p}$$

ونجد أنّ $E[g(T)]=0$ من أجل جميع قيم P وهذا يعني أنّه:

$$\sum_{t=0}^{n} g(t).\binom{n}{t}.\rho^t.(1-p)^n=0$$

وبما أنّ $(1-p)^n\neq 0$ نجد أنّ:

$$\sum_{t=0}^{n} g(t).\binom{n}{t}.\rho^t=0 \quad \text{من أجل جميع قيم } \rho$$

وبما أنّ الطرف الأيسر من المطابقة الأخيرة كثير حدود في ρ فإنّ جميع الأمثال يجب أنّ تساوي الصفر أي أنّ:

$g(t)\binom{n}{t}\equiv 0$ من أجل جميع قيم t وبالتالي $g(t)\equiv 0$ وذلك أياً كانت t وحسب التعريف يكون T إحصاءً تاماً.

مثال: لتكن $X_1,X_2,\dots,X_n$ عينة عشوائية حجمها n مأخوذة من المجتمع الإحصائي الموصوف بدالة الكثافة الاحتمالية : $f_X(x;\alpha,\lambda)=\frac{\alpha^\lambda}{\Gamma(\lambda)}x^{\lambda-1}e^{-\alpha x}$, وبفرض أنّ $\alpha=\frac{1}{\beta}$ و λ معلومة والمطلوب:

1. أوجد مقدِّر للوسيط β بطريقة الاحتمالية العظمى.
2. أوجد مقدِّر للوسيط β بطريقة العزوم .
3. هل المقدِّر الناتج غير منحاز بالنسبة β وهل هو متّسق .
4. هل المقدِّر الناتج غير منحاز ذو تباين أصغري.
5. بفرض أنّ $T=\sum_{i=1}^{n} X_i$ معلوم , أثبت أنّ T هو احصاء كافي للوسيط β وهل هو احصاء تام.

الحل: **(طريقة اولى لحل التمرين)**

.1

$$L(x_1,\dots,x_n;\beta)=\frac{1}{[\beta^\lambda\Gamma(\lambda)]^n}\prod_{i=1}^{n}(x_i)^{\lambda-1}e^{-\frac{1}{\beta}\sum_{i=1}^{n}x_i}$$

$$lnL(x_1,\dots,x_n;\beta)=-n\lambda\, ln\beta-n\ln\Gamma(\lambda)+(\lambda-1)\sum_{i=1}^{n} lnx_i-\frac{1}{\beta}\sum_{i=1}^{n}x_i$$

$$\frac{\partial lnL(x_1,\dots,x_n;\beta)}{\partial\beta}=\frac{-n\lambda}{\beta}+\frac{1}{\beta^2}\sum_{i=1}^{n}x_i$$

ومن معادلة الاحتمالية العظمى $\left.\frac{\partial \ln L(x_1,\dots,x_n;\beta)}{\partial\beta}\right|_{\beta=\hat{\beta}}=0$ نجد :

$$\frac{n\lambda}{\hat{\beta}}=\frac{1}{\hat{\beta}^2}\sum_{i=1}^{n}x_i \Rightarrow \hat{\beta}=\frac{1}{\lambda}\bar{X}$$

وهو مقدِّر الاحتمالية العظمى للوسيط β .

.2

$$EX|_{\beta=\hat{\beta}}=\frac{1}{n}\sum_{i=1}^{n}x_i \Rightarrow \frac{\lambda}{\hat{\alpha}}=\bar{X} \Rightarrow \lambda\hat{\beta}=\bar{X} \Rightarrow \hat{\beta}=\frac{1}{\lambda}\bar{X}$$

وهو مقدِّر العزوم للوسيط β .

.3

$$E(\hat{\beta})=E\left(\frac{1}{\lambda}\bar{X}\right)=\frac{1}{\lambda}E(\bar{X})=\frac{1}{\lambda}E(X)=\frac{1}{\lambda}\cdot\frac{\lambda}{\alpha}=\frac{1}{\alpha}=\beta$$

وبالتالي فإنّ $\hat{\beta}$ غير منحاز بالنسبة لـ β .

$$Var(\hat{\beta})=Var\left(\frac{1}{\lambda}\bar{X}\right)=\frac{1}{\lambda^2}Var(\bar{X})=\frac{1}{\lambda^2}\frac{Var(X)}{n}=\frac{1}{\lambda^2 n}\cdot\frac{\lambda}{\alpha^2}=$$

$$=\frac{1}{n\lambda\alpha^2}=\frac{\beta^2}{\lambda n}\xrightarrow[n\to\infty]{}0$$

اذا $\hat{\beta}$ متّسق .

.4

$$\frac{\partial^2 \ln L(x_1,\dots,x_n;\beta)}{\partial\beta^2}=\frac{n\lambda}{\beta^2}-\frac{2}{\beta^3}\sum_{i=1}^{n}x_i$$

$$\Rightarrow E\left[-\frac{\partial^2 \ln L(x_1,\dots,x_n;\beta)}{\partial\beta^2}\right]=-\frac{n\lambda}{\beta^2}+\frac{2}{\beta^3}E\left[\sum_{i=1}^{n}x_i\right]$$

$$=-\frac{n\lambda}{\beta^2}+\frac{2}{\beta^3}E[n\bar{X}]=-\frac{n\lambda}{\beta^2}+\frac{2n}{\beta^3}E[X] \qquad ;\left(E[X]=\frac{\lambda}{\alpha}=\beta\right)$$

$$=-\frac{n\lambda}{\beta^2}+\frac{2n\lambda}{\beta^2}=\frac{n\lambda}{\beta^2}$$

نلاحظ أنّ : $Var(\hat{\beta})=\frac{\beta^2}{\lambda n}=\frac{1}{\frac{\lambda n}{\beta^2}}=\frac{\left[\left(E(\hat{\beta})\right)'\right]^2}{E\left[-\frac{\partial^2 \ln L(x_1,\dots,x_n)}{\partial\beta^2}\right]}$

اذا المساواة محققة في متباينة كرامر ـ راو وبالتالي $\hat{\beta}$ غير منحاز ذو تباين اصغري بالنسبة للوسيط β .

طريقة ثانية:

$$L(x_1,\dots,x_n;\beta)=\left(\beta^{-n\lambda}\right)\left(\Gamma(\lambda)\right)^{-n}\left(\prod_{i=1}^{n}(x_i)^{\lambda-1}\right)e^{-\frac{1}{\beta}\sum_{i=1}^{n}x_i}$$

$$;\qquad \left(\frac{1}{\beta}\sum_{i=1}^{n}x_i=\frac{n}{\beta}\cdot\frac{\sum_{i=1}^{n}x_i}{n}\cdot\frac{\lambda}{\lambda}=\frac{n}{\beta}\cdot\bar{X}\cdot\frac{\lambda}{\lambda}\right)$$

$$=\underbrace{\frac{1}{[\Gamma(\lambda)]^n}\left(\prod_{i=1}^{n}(x_i)^{\lambda-1}\right)}_{L_1(x_1,\dots,x_n)}\cdot\underbrace{e^{\left(-\frac{n\lambda}{\beta}\right)\left(\frac{1}{\lambda}\bar{X}\right)-n\lambda\ln\beta}}_{e^{\beta_1\hat{\beta}+\beta_2}}=L_1(x_1,\dots,x_n)\cdot e^{\beta_1\hat{\beta}+\beta_2}$$

حيث : $\beta_2 = -n\lambda\ ln\beta$ ، $\beta_1 = -\frac{n\lambda}{\beta}$ توابع لـβ فقط

$\{\hat{\beta} = \frac{1}{\lambda}\bar{X}$ لـ مقدِّر β $\{L_1(x_1, \dots, x_n)$ تابع لمتغيرات العينة فقط

إذا $\hat{\beta}$ غير منحاز بالنسبة لـ β .

$$E(\hat{\beta}) = -\frac{\beta_2'}{\beta_1'} = \frac{(-n\lambda\ ln\beta)'}{\left(-\frac{n\lambda}{\beta}\right)'} = \frac{-\frac{n\lambda}{\beta}}{-\frac{n\lambda}{\beta^2}} = \beta$$

$$Var(\hat{\beta}) = \frac{\beta_1''\beta_2' - \beta_2''\beta_1'}{(\beta_1')^3} = \frac{\left(-\frac{2n\lambda}{\beta^3}\right)\left(-\frac{n\lambda}{\beta}\right) - \left(\frac{n\lambda}{\beta^2}\right)\left(\frac{n\lambda}{\beta^2}\right)}{\left(\frac{n\lambda}{\beta^2}\right)^3} = \frac{\frac{1}{\beta^4}(2n^2\lambda^2 - n^2\lambda^2)}{\frac{n^3\lambda^3}{\beta^6}} = \frac{\beta^2}{n\lambda} \xrightarrow[n\to\infty]{} 0$$

اذا $\hat{\beta}$ متّسق .

.5

$$L(x_1, \dots, x_n; \beta) = \frac{\beta^{-n\lambda}}{[\Gamma(\lambda)]^n}\left(\prod_{i=1}^n (x_i)^{\lambda-1}\right)e^{-\frac{1}{\beta}\sum_{i=1}^n x_i}$$

$$= \underbrace{\left[\frac{\beta^{-n\lambda}}{\Gamma(n\lambda)}\left(\sum_{i=1}^n x_i\right)^{n\lambda-1} e^{-\frac{1}{\beta}\sum_{i=1}^n x_i}\right]}_{g_T(t;\beta)} \underbrace{\left[\frac{\Gamma(n\lambda)}{[\Gamma(\lambda)]^n} \cdot \frac{\left(\prod_{i=1}^n x_i\right)^{\lambda-1}}{\left(\sum_{i=1}^n x_i\right)^{n\lambda-1}}\right]}_{H(x_1,\dots,x_n)}$$

$$= \underbrace{g_T(t;\beta)}_{\text{دالة كثافة } T} \cdot \underbrace{H(x_1, \dots, x_n)}_{\text{دالة تتعلق بمتغيرات العينة فقط}}$$

إذا T إحصاء كافي للوسيط β .

بما أنّ T هو إحصاء كافي فلإثبات أنّه تام يكفي إثبات أنّ تحقق الشرط $E[h(T)] \equiv 0$ يودي إلى تحقق الشرط $h(T) \equiv 0$.

$$E[h(T)] = \int_0^{+\infty} h(t)\, g_T(t;\beta)dt = \frac{\beta^{-n\lambda}}{\Gamma(n\lambda)}\int_0^{+\infty} h(t)\, t^{n\lambda-1}\, e^{-\frac{1}{\beta}t}dt$$

$$E[h(T)] \equiv 0 \Leftrightarrow \int_0^{+\infty} h(t)\, \underbrace{t^{n\lambda-1}}_{\text{كثير حدود}}\ \underbrace{e^{-\frac{1}{\beta}t}}_{>0}\, dt = 0$$

$$\Leftrightarrow h(t)\, t^{n\lambda-1}\, e^{-\frac{1}{\beta}t}dt \equiv 0 \qquad \Leftrightarrow h(t) \equiv 0 \quad ;\forall\, t$$

إذا T هو إحصاء تام .

الفصل الثالث

التقدير بفترة

Interval estimation

لقد قدمنا في الفقرة السابقة فكرة بسيطة عن التقدير النقطي وعن بعض أهم المقدِّرات النقطية ، ثم شرحنا الطرائق الهامة في تقدير الوسطاء الإحصائية، ولنفرض الآن أنّنا حصلنا على مقدِّر نقطي لوسيط ما ، وكان ، مثلاً غير منحاز وذا تباين أصغري ، فإنّ القيمة التقديرية الناتجة تعطينا فكرة عن مقدار الوسيط الحقيقي . ولكن ما من سبب يدعونا إلى التوقع إلى أنّ المقدِّر النقطي يساوي تماماً الوسيط المجهول ، وربما كان من المرغوب فيه أنّ ننشئ فترة نتوقع أنّ نجد فيها قيمة هذا الوسيط المجهول ، بحيث يمثل طرفا هذه الفترة عبارة عن إحصاءين L_1 وL_2 تابعين في متغيرات العينة العشوائية المختارة بحيث يكون احتمال أنّ يقع الوسيط الحقيقي θ بين L_1 وL_2 يساوي $1-\alpha$ ونكتب : $P\left(L_1 \leq \theta \leq L_2\right) = 1-\alpha$

وهذا هو معيار الثقة الذي يمكن أنّ يتمتع به مثل هذا المقدِّر .

تسمى الفترة $\left[L_1 , L_2\right]$ بفترة ثقة موافقة لـ $100(1-\alpha)\%$ للوسيط المجهول θ ، حيث يسمى الاحتمال $1-\alpha$ بأمثال الثقة ، كما وتسمى α بمستوى الثقة أو مستوى الدلالة أو مستوى المعنوية .

لإيجاد فترة الثقة المطلوبة نتبع الخطوات الآتية :

1- نفتش عن كمية تكون تابعة في متغيرات العينة وتحوي الوسيط المجهول ويكون توزيعها هو أحد التوزيعات الشهيرة . تسمى هذه الكمية بالكمية المحورية .

2- نحصر هذه الكمية المحورية بين قيمتين معياريتين مأخوذتين من توزيع العينة بحيث يكون احتمال أنّ تقع هذه الكمية المحورية محصورة بين هاتين القيمتين المعياريتين يساوي $1-\alpha$.

3- نعزل الوسيط المجهول ونوجد فترة الثقة المطلوبة .

1 - فترة الثقة لمتوسط مجتمع طبيعي μ .

مبرهنة (1) : لتكن $X_1 , X_2 ,..., X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي متوسطه μ وتباينه σ^2 معلوم . أنّ $100(1-\alpha)\%$ فترة ثقة للمتوسط μ تعطى بالشكل :

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \;,\; \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

حيث n هو حجم العينة ، $z_{1-\frac{\alpha}{2}}$ هو العدد الحقيقي الذي يحصر على يساره مساحة قدرها $\left(1-\frac{\alpha}{2}\right)$ من منحني التوزيع الطبيعي المعياري . أو هو المئين $\left(1-\frac{\alpha}{2}\right)$ للتوزيع الطبيعي المعياري .

البرهان :

بما أنّ X له التوزيع الطبيعي ، فإنّ الكمية $Z=\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sigma}$ له التوزيع الطبيعي المعياري وذلك اعتماداً على نظرية النهاية المركزية . كما ونلاحظ أنّ هذه الكمية تحوي متغيرات العينة بالإضافة إلى الوسيط μ المجهول، وبالتالي فإنّ هذه الكمية تصلح لأنّ تكون هي الكمية المحورية المناسبة . نحصر هذه الكمية المحورية بين القيمتين المعياريتين للتوزيع الطبيعي المعياري $z_{1-\frac{\alpha}{2}}$ ، $-z_{1-\frac{\alpha}{2}}$ ولنثبت أنّ :

$$P\left(-z_{1-\frac{\alpha}{2}} \le Z \le z_{1-\frac{\alpha}{2}}\right)=1-\alpha \qquad (1)$$

سوف ننطلق من العلاقة :

$$P\left(Z \le z_{\alpha}\right)=\alpha$$

نأخذ الطرف الأيسر فنجد :

$$\begin{aligned}
P\left(-z_{1-\frac{\alpha}{2}} \le Z \le z_{1-\frac{\alpha}{2}}\right) &= P\left(Z \le z_{1-\frac{\alpha}{2}}\right)-P\left(Z \le -z_{1-\frac{\alpha}{2}}\right)\\
&= P\left(Z \le z_{1-\frac{\alpha}{2}}\right)-P\left(Z \ge z_{1-\frac{\alpha}{2}}\right)\\
&= P\left(Z \le z_{1-\frac{\alpha}{2}}\right)-\left[1-P\left(Z \le z_{1-\frac{\alpha}{2}}\right)\right]\\
&= 2\,P\left(Z \le z_{1-\frac{\alpha}{2}}\right)-1\\
&= 2\left(1-\frac{\alpha}{2}\right)-1\\
&= 2-\alpha-1=1-\alpha
\end{aligned}$$

نعوض Z بما يساويها في العلاقة (1) فنجد :

$$P\left(-z_{1-\frac{\alpha}{2}} \le \frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sigma} \le z_{1-\frac{\alpha}{2}}\right)=1-\alpha$$

بالإصلاح نجد :

$$P\left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \Rightarrow$$

$$P\left(L_1 \le \mu \le L_2\right) = 1 - \alpha$$

وبالتالي فإنّنا نستطيع كتابة فترة الثقة بالشكل : $\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad , \quad \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$

ملاحظة :

1 – أنّ طول فترة الثقة هو عبارة عن الفرق بين الحدين الأدنى والأعلى للفئة , فإذا فرضنا أنّ طول فترة الثقة هو a , فإنّ طول فترة الثقة في هذه الحالة يعطى بالشكل : $a = 2 \; z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

2 – نلاحظ أنّه كلما ازداد مستوى الثقة , كلما ازداد طول لفترة وأنّ طول فترة الثقة يقل بزيادة حجم العينة .

مثال : أخذت عينة عشوائية حجمها $n = 100$ ، من مجتمع طبيعي ، تباينه $\sigma^2 = 25$. فإذا كان متوسط هذه العينة هو 56.18 . أوجد % 95 فترة ثقة لمتوسط هذا المجتمع الطبيعي μ .

الحل :

لدينا : $1 - \alpha = 0.05 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow$

$$1 - \frac{\alpha}{2} = 0.975 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

أنّ فترة الثقة تعطى بالشكل :

$$C.I = \left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \quad ; \quad \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

$$= \left[56.18 - \left(\frac{5}{10}\right)(1.96) \quad ; \quad 56.18 + \left(\frac{5}{10}\right)(1.96) \right]$$

$$= \left[55.20 \quad ; \quad 57.16 \right]$$

مبرهنة (2) : لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي متوسطه μ وتباينه σ^2 مجهولان . أنّ $100(1-\alpha)\%$ فترة الثقة للمتوسط μ يعطى بالشكل :

$$\left[\overline{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \; ; \; \overline{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \right]$$

حيث n هو حجم العينة ، $t_{\left(1-\frac{\alpha}{2}\right)}(n-1)$ هو المئين $100\left(1 - \frac{\alpha}{2}\right)$ للتوزيع t بعدد n−1 درجة من الحرية أو هو العدد الحقيقي الذي يحصر على يساره مساحة قدرها $1 - \frac{\alpha}{2}$ من منحني توزيع ستيودنت درجة حريته $n - 1$.

البرهان :

نلاحظ أنّ الكمية المحورية السابقة Z لا تصلح لأنّ تكون كمية محورية لهذه الحالة ، وذلك بسبب وجود الوسيط σ المجهول ، لذا نفتش عن الكمية المحورية المناسبة .

إنّ الكمية المحورية هنا هي $T = \frac{\sqrt{n}\left(\overline{X}-\mu\right)}{S}$ وهي دالة في الإحصاءين $\overline{X}$ و S والوسيط μ ، وكما نعلم فإنّ توزيعها الاحتمالي هو توزيع ستيودنت درجة حريته $n-1$ وهذا التوزيع مستقل عن μ . وبطريقة مشابهة لما أثبتناه في المبرهنة السابقة نثبت أنّ :

$$P\left(-t_{1-\frac{\alpha}{2}}(n-1) \le \frac{\sqrt{n}\left(\overline{X}-\mu\right)}{S} \le t_{1-\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

بالإصلاح نجد أنّ : $P\left(\overline{X}-\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1) \le \mu \le \overline{X}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$

وبالتالي فإنّ الفترة المطلوبة هي : $\left[\overline{X}-\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\ ;\ \overline{X}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\right]$

مثال: عينة عشوائية حجمها $n = 26$ مأخوذة من مجتمع طبيعي وسيطاه μ و σ^2 . فإذا كان متوسط هذه العينة $\overline{X} = 32.8$ وأنّ انحرافها المعياري $S = 4.51$ فأوجد % 95 فترة ثقة للوسيط μ .

الحل :

بما أنّ μ و σ^2 مجهولتان ، فإنّ فترة الثقة المطلوبة هي:

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\ ;\ \overline{X}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\right]$$

لدينا $\overline{X} = 32.8$ ، $S = 4.51$ ، $n = 26$. ولدينا $1-\alpha = 95\% \Rightarrow \alpha = 0.05$

$$\Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow 1-\frac{\alpha}{2} = 0.075 \Rightarrow t_{1-\frac{\alpha}{2}}(n-1) = t_{0.975}(25) = 2.06$$

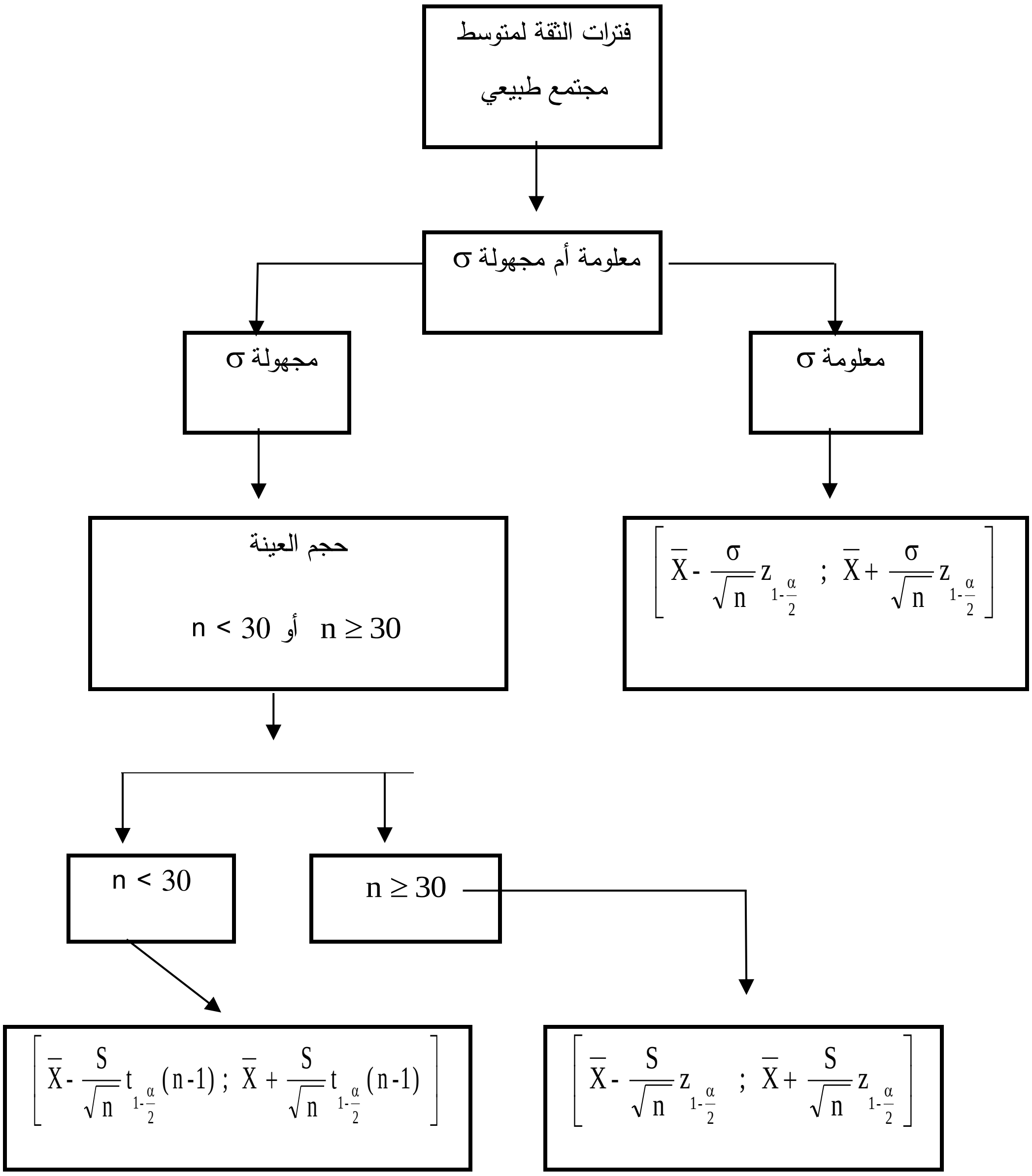
بالتعويض نجد أنّ فترة الثقة لـ μ هي : $[\,30.95\ ,\ 34.65\,]$.

ملاحظة: إذا كانت σ^2 مجهولة , وكانت $n \ge 30$, فإنّ فترة الثقة لمتوسط مجتمع طبيعي μ , تعطى بالشكل:

$$\left[\overline{X}-z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\ ,\ \overline{X}+z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

حيث استبدلنا الانحراف المعياري للمجتمع σ بالانحراف المعياري للعينة S في المبرهنة الأولى .

سوف نوضح فيما يلي مخططاً في إيجاد فترات الثقة لمتوسط مجتمع طبيعي :



2 – فترة الثقة للنسبة p في التوزيع الحداني $b(x\,;1\,,p)$:

لاحظنا في الفصل السابق أنّه إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية حجمها n مأخوذة من التوزيع الحداني $b(x\,;1\,,p)$ ، فإنّ التقدير النقطي $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ هو وسط حسابي وأنّ الوسط الحسابي للتوزيع $\mu = p$ ، ومن ثم فإنّه واعتماداً على نظرية النهاية المركزية نجد أنّ توزيع $\hat{p}$ له التوزيع الطبيعي $N\left(p, \frac{p(1-p)}{n}\right)$ ، ومن ثَمَّ فإنّ توزيع الكمية $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ هو التوزيع الطبيعي المعياري ، وبما أنّ p مجهولة ، لذا نضطر إلى تقريب p بـ $\hat{p}$ وبالتالي فإنّ الكمية $Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ لها التوزيع الطبيعي المعياري .

مبرهنة (3): لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي يخضع للتوزيع الحداني b (x ; 1 , p). أنّ $100(1-\alpha)\%$ فترة ثقة للنسبة p هي :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad ; \quad \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

حيث حجم العينة n كبيراً وأنّ $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

البرهان: نلاحظ أنّ الكمية $Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ هي الكمية المحورية المناسبة في هذه الحالة ، لذا نقوم بحصرها بين القيمتين المعياريتين $z_{1-\frac{\alpha}{2}}$ ، $-z_{1-\frac{\alpha}{2}}$ ثم نعزل الوسيط المجهول ونحصل على فترة الثقة المطلوبة .

مثال: لنفترض أنّنا نرغب في إجراء مسح إحصائي لتقدير نسبة الذكور في مدينة حمص ، فإذا أخذت عينة عشوائية مؤلفة من 400 شخص ، ووجدنا أنّ عدد الذكور هو 180 شخصاً فأوجد % 95 فترة ثقة لنسبة الذكور p .

الحل : لدينا: $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.25 \Rightarrow 1-\frac{\alpha}{2} = 0.975 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_1 = \frac{180}{400} = 0.45$$

وبالتالي فإنّ فترة الثقة المطلوبة هي :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad ; \quad \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] =$$

$$= \left[0.45 - 1.96 \sqrt{\frac{(0.45)(0.55)}{400}} \quad ; \quad 0.45 + 1.96 \sqrt{\frac{(0.45)(0.55)}{400}} \right]$$

$$= [0.45 - 0.049 \quad ; \quad 0.45 + 0.049] = [0.4010 ; 0.499]$$

3 – فترة ثقة لتباين مجتمع طبيعي :

وجدنا سابقاً أنّ تقدير التباين ضروري عند القيام باستقراء حول المتوسط ، حيث يعد التباين σ^2 الهدف الرئيسي للتحريات التجريبية ، وذلك في العديد من المسائل التطبيقية . وكما نعلم فإنّ أجهزة القياسات العلمية يجب أنّ تقدم قياسات غير منحازة من جهة وبخطأ قياس صغير جداً من جهة أخرى . فمقياس درجة الحرارة سوف لن يكون له أية فائدة إذا كان الانحراف المعياري لخطأ القياس كبيراً ، وفي الحقيقة نتمكن ، على الأغلب ، من تصحيح الانحياز في القياس ، ولكن دقة القياس تكون مقاسة بحجم الانحراف المعياري لخطأ القياس هو في

العادة تابع لتصميم الجهاز نفسه ولا يمكن التحكم به . وبشكل عام نرغب عادة في المحافظة على أصغر تباين ممكن في قياسات الميزات النوعية لإنتاج صناعي وذلك لإنجاز نوع من التحكم في الطريقة الصناعية وجعل النسبة المئوية للمنتجات ذات النوعية الفقيرة أصغر ما يمكن .

مبرهنة (4) : لتكن $X_1 , X_2 ,..., X_n$ عينة عشوائية مأخوذة من مجتمع له التوزيع الطبيعي وسيطاه μ , σ^2 ، حيث $\mu = \mu_o$ معلومة . أنّ $100(1-\alpha)\%$ فترة ثقة لـ σ^2 تعطى بالشكل : $\left[\dfrac{n S^2}{\chi^2_{\frac{\alpha}{2}}(n)} \; ; \; \dfrac{n S^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \right]$

حيث يرمز $\chi^2_{\alpha}(n)$ للمئين $100.\alpha$ لتوزيع كاي مربع درجة حريته n .

S^2 هو تباين العينة ويعطى بالشكل : $S^2 = \dfrac{1}{n} \sum\limits_{i=1}^{n} (X - \mu_o)^2$

البرهان : أنّ الكمية المحورية المطلوبة في هذه الحالة هي $\dfrac{n S^2}{\sigma^2}$ ، ونعلم أنّ توزيعها هو توزيع كاي مربع درجة حريته (n) ، وهو توزيع مستقل عن σ^2، وبالتالي نحصر هذه الكمية المحورية بين القيمتين المعياريتين $\chi^2_{1-\frac{\alpha}{2}}(n)$ و $\chi^2_{\frac{\alpha}{2}}(n)$ ثم نعزل الوسيط σ^2 المجهول ، فنحصل على فترة الثقة المطلوبة .

مبرهنة (5) : لتكن $X_1 , X_2 ,..., X_n$ عينة عشوائية مأخوذة من توزيع طبيعي وسيطاه μ , σ^2 مجهولان . أنّ $100(1-\alpha)\%$ فترة ثقة للتباين σ^2 تعطى بالشكل : $\left[\dfrac{(n-1)\, S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \; ; \; \dfrac{(n-1)\, S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \right]$

حيث يرمز $\chi^2_{\alpha}(n-1)$ للمئين $100.\alpha$ لتوزيع كاي مربع درجة حريته (n – 1) و S^2 هو تباين العينة ويعطى بالعلاقة : $S^2 = \dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \overline{X} \right)^2$

البرهان : أنّ الكمية المحورية المطلوبة في هذه الحالة هي $(n-1)\dfrac{S^2}{\sigma^2}$ ، ونعلم أنّ توزيعها هو توزيع كاي مربع درجة حريته (n – 1) ، وهو توزيع مستقل عن σ^2، وبالتالي نكتب :

$$P\left[\chi^2_{\frac{\alpha}{2}}(n-1) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n-1) \right] = 1-\alpha$$

بالإصلاح نجد : $$P\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \right] = 1-\alpha$$

مثال: أخذت عينة عشوائية حجمها $n = 10$ من المجتمع الطبيعي $N(\mu, \sigma^2)$ ، ووجد أنّ تباينها $S^2 = 117.12$.

أوجد 95 % فترة ثقة للتباين σ^2 .

الحل : لدينا : $n = 10 \quad , \quad n-1 = 9$

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow$$

$$1-\frac{\alpha}{2} = 0.975 \Rightarrow \chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0.025}(9) = 2.7 \ , \ \chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0.975}(9) = 19.02$$

أنّ فترة الثقة المطلوبة تعطى بالشكل : $\left[\dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \ ; \ \dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \right] = [55.42 \ ; \ 390.40]$

4 – فترة الثقة للفرق بين متوسطي مجتمعين طبيعيين :

نرغب في كثير من الأحيان في تقدير الفرق بين متوسطي مجتمعين طبيعيين فقد نرغب في مقارنة معدلي الإنتاج في شركة الصناعات الدوائية عند استخدام المواد الأولية الواردة من اثنين من الممولين ، فنأخذ عينة من الإنتاج الأسبوعي في كل حالة ونستخدم المعلومات التي تقدمها هاتان العينتان للقيام باستقراء يتعلق حول معدلي الإنتاج ، أو قد ترغب شركة صناعة الإطارات أنّ تقدّر الفرق في مواصفات تلبيس نوعين من الإطارات .

نفترض في كل من هذين المثالين وجود مجتمعين إحصائيين متمايزين متوسط الأول μ_1 وتباينه σ_1^2 معلوم ومتوسط الثاني μ_2 وتباينه σ_2^2 معلوم . فإذا سحبنا من هذين المجتمعين عينتين عشوائيتين مستقلتين حجماهما m و n على الترتيب ، فإنّ التقدير النقطي للفرق بين متوسطي المجتمعين $(\mu_1 - \mu_2)$ هو $(\overline{X} - \overline{Y})$ ومن ثمَّ فإنّ توزيع الفرق $(\overline{X} - \overline{Y})$ هو التوزيع الطبيعي متوسطه $(\mu_1 - \mu_2)$ وتباينه $\dfrac{\sigma_2^2}{n} + \dfrac{\sigma_1^2}{m}$ ، وبالتالي فإنّ توزيع المتغير $Z = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}$ هو التوزيع الطبيعي المعياري .

مبرهنة(6): لتكن $X_1, X_2, \dots, X_m$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N(\mu_1, \sigma_1^2)$ ولتكن $Y_1, Y_2, \dots, Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N(\mu_1 \ , \ \sigma_1^2)$ ، ولنفرض أنّ تباينيهما معلومان .

أنّ $100(1-\alpha)\%$ فترة ثقة للفرق $(\mu_1 - \mu_2)$ تعطى بالشكل $[L_1, L_2]$ حيث :

$$L_1 = (\overline{X} - \overline{Y}) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \quad , \quad L_2 = (\overline{X} + \overline{Y}) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

البرهان :أنّ الكمية المحورية المناسبة لهذه الحالة هي : $Z=\dfrac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}$

وأنّ توزيعها هو التوزيع الطبيعي المعياري ، لذلك نحصر هذه الكمية بين القيمتين المعياريتين $z_{1-\frac{\alpha}{2}}$ ، $-z_{1-\frac{\alpha}{2}}$

ثم نعزل الوسيط المجهول $(\mu_1-\mu_2)$ فنحصل على فترة الثقة المطلوبة .

مثال: إذا كان متوسط عمر صمام جهاز إلكتروني من نوع ألفا هو 6.5 سنة بتباين وقدره 0.81 سنة بينما متوسط عمر صمام في جهاز إلكتروني آخر من نوع بيتا هو 6 سنة بتباين وقدره 0.64 سنة. أخذنا عينة حجمها $m=36$ من مجتمع صمامات النوع الأول ، ثم عينة أخرى مستقلة عن الأولى حجمها $n=49$ من مجتمع صمامات النوع الثاني ، فإذا فرضنا أنّ مجتمعي الصمامات هما طبيعيان ، فأوجد الاحتمال الآتي: $P\left(\overline{X}-\overline{Y}\ge 1\right)$.

الحل : لنوجد التوقع الرياضي والتباين للمتغير $\left(\overline{X}-\overline{Y}\right)$ كما يلي : $E\left(\overline{X}-\overline{Y}\right)=\mu_1-\mu_2=6.5-6=0.5$

وكذلك (لأنّ العينتين مستقلتان) ، $Var\left(\overline{X}-\overline{Y}\right)=\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}=\dfrac{0.81}{36}+\dfrac{0.64}{49}=(0.189)^2$

وبالتالي نجد :

$$P\left(\overline{X}-\overline{Y}\ge 1\right)=P\left(Z\ge\frac{1-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}\right)$$

$$=P\left(Z\ge\frac{1-0.5}{0.189}\right)=p(Z\ge 2.65)=1-P(Z\le 2.65)=1-0.996=0.004$$

مبرهنة (7): لتكن $X_1,X_2,\ldots,X_m$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N\left(\mu_1\ ,\ \sigma_1^2\right)$ ، ولتكن $Y_1,Y_2,\ldots,Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N\left(\mu_2\ ,\ \sigma^2\right)$ ، ولنفرض أنّ σ^2 مجهولة . أنّ $100(1-\alpha)\%$ فترة ثقة للفرق $(\mu_1-\mu_2)$ تعطى بالشكل $[L_1,L_2]$ حيث :

$$L_1=\left(\overline{X}-\overline{Y}\right)-t_{1-\frac{\alpha}{2}}(m+n-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}}$$

$$L_2=\left(\overline{X}-\overline{Y}\right)+t_{1-\frac{\alpha}{2}}(m+n-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}}$$

حيث S_p^2 هو التباين المكافئ ويعطى بالعلاقة : $S_p^2=\dfrac{1}{m+n-2}\left[\sum_{i=1}^{m}\left(X_i-\overline{X}\right)^2+\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2\right]$

البرهان : نحن نعلم من الفصل السابق أنّ توزيع الكمية : $T=\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_P\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$

هو توزيع ستيودنت درجة حريته $(m+n-2)$ ، ونلاحظ أنّها تحوي مقادير معلومة وتحوي الوسيط المجهول ، لذا فهي تصلح لأنّ تكون الكمية المحورية المناسبة لهذه الحالة . وبحصر هذه الكمية T بين القيمتين المعياريتين $t_{1-\frac{\alpha}{2}}(m+n-2)$ وَ $-t_{1-\frac{\alpha}{2}}(m+n-2)$ وعزل الوسيط المجهول $(\mu_1-\mu_2)$ ، نحصل على $100(1-\alpha)\%$ فترة ثقة للفرق $(\mu_1-\mu_2)$ المطلوبة .

مثال: إذا كانت 41 , 39 , 31 , 29 , 35 , 23 , 20 عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu_2\ ,\ \sigma^2)$ وكانت 25 , 45 , 37 , 43 , 34 عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N(\mu_2\ ,\ \sigma^2)$. أوجد 95 % فترة ثقة للفرق $(\mu_1-\mu_2)$.

الحل : نلاحظ أنّ : $1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \dfrac{\alpha}{2}=0.025 \Rightarrow 1-\dfrac{\alpha}{2}=0.975$

وأنّ : $m=7\ ,\ n=5\ ,\ m+n-2=10$ وبالتالي فإنّ القيمة المعيارية لتوزيع ستيودنت درجة حريته (10) هي : $t_{1-\frac{\alpha}{2}}(m+n-2)=t_{0.975}(10)=2.228$.

لنوجد الآن متغيرات العينتين : $\overline{X}=31.1\ ,\ S_1^2=61.5\ ,\ \overline{Y}=36.8\ ,\ S_2^2=63.2$

وبالتالي فإنّ التباين المكافئ يعطى بالشكل :

$$S_P^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}=\frac{7(61.5)+5(63.2)}{10}=62.17 \Rightarrow S_P=7.88$$

وبالتالي فإنّ حدي الثقة المطلوبان هما :

$$L_1=\left(\overline{X}-\overline{Y}\right)-t_{1-\frac{\alpha}{2}}(m+2-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}}=-5.7-(2.228)(7.88)(0.585)=-15.97$$

$$L_2=\left(\overline{X}-\overline{Y}\right)+t_{1-\frac{\alpha}{2}}(m+n-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}}=-5.7+(2.228)(7.88)(0.585)=4.57$$

5 – فترة ثقة للفرق (p_1-p_2) :

مبرهنة (8): لتكن $X_1,X_2,\dots,X_m$ عينة عشوائية مأخوذة من مجتمع إحصائي له توزيع حداني من الشكل $b(x\ ;1,p_1)$ ، ولتكن $Y_1,Y_2,\dots,Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من مجتمع إحصائي له توزيع حداني من الشكل $b(x;1,p_2)$. أنّ $100(1-\alpha)\%$ فترة ثقة للفرق (p_1-p_2) تعطى بالشكل $[L_1,L_2]$ حيث :

$$L_1 = \left(\hat{p}_1 - \hat{p}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{m} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n}}$$

$$L_2 = \left(\hat{p}_1 - \hat{p}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{m} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n}}$$

مثال: يرمي كل من عمر و مُزن السهام على هدف ، فأصابه مُزن 64 مرة ، وأصابه عمر 56 مرة . بفرض أنّ p_1 تمثل النسبة الصحيحة لإصابة مُزن للهدف ، وأنّ p_2 تمثل النسبة الصحيحة لإصابة عمر للهدف . أوجد 95 % فترة ثقة للفرق $\left(p_1 - p_2\right)$.

الحل: لدينا : $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \dfrac{\alpha}{2} = 0.025 \Rightarrow 1-\dfrac{\alpha}{2} = 0.975 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

$$\hat{p}_1 = \frac{64}{100} = 0.64 \quad ; \quad \hat{p}_2 = \frac{56}{100} = 0.56$$

أنّ فترة الثقة تعطى بالشكل :

$$C.I. = \left(\hat{p}_1 - \hat{p}_2\right) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{m} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n}}$$

$$= \left(0.64 - 0.56\right) \pm 1.96 \sqrt{\frac{(0.64)\,(0.36)}{100} + \frac{(0.56)\,(0.44)}{100}}$$

$$= 0.08 \pm 1.96\sqrt{0.048+0.049}$$

$$= 0.08 \pm 0.61$$

وبالتالي فإنّ فترة الثقة المطلوبة هي : $[\,-0.53 \;\; ; \;\; 0.60\,]$.

6– فترة ثقة لنسبة تبايني مجتمعين طبيعيين :

مبرهنة(9): لتكن $X_1, X_2, \ldots, X_m$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي $N\left(\mu_1, \sigma_1^2\right)$ ولتكن $Y_1, Y_2, \ldots, Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من مجتمع إحصائي آخر له التوزيع الطبيعي $N\left(\mu_2\,,\,\sigma_2^2\right)$. أنّ $100(1-\alpha)\%$ فترة ثقة للنسبة $\dfrac{\sigma_2^2}{\sigma_1^2}$ هي :

$$\left[\frac{S_2^2}{S_1^2} f_{\frac{\alpha}{2}}(m-1\,,\,n-1) \;\; ; \;\; \frac{S_2^2}{S_1^2} f_{1-\frac{\alpha}{2}}(m-1\,,\,n-1)\right]$$

البرهان:

نحن نعلم أنّ الكمية $F = \dfrac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}$ تتوزع وفق التوزيع F بدرجتي حرية $(m - 1)$ و $(n - 1)$ على الترتيب ، كما

ونلاحظ أنّ هذه الكمية تحوي الوسيط المجهول وتحوي متغيرات العينة ، لذا فإنّ هذه الكمية تصلح لأنّ تكون الكمية المحورية المناسبة في هذه الحالة ، ويحصرها بين القيمتين المعياريتين $f_{\frac{\alpha}{2}}(m-1, n-1)$ وَ $f_{1-\frac{\alpha}{2}}(m-1, n-1)$ وبعزل الوسيط المجهول نحصل على فترة الثقة المطلوبة .

مثال: أخذت عينة حجمها $m = 6$ ، وتباينها $S_1^2 = 80$ ، من توزيع طبيعي $N\left(\mu_1 , \sigma_1^2\right)$ ، ثم أخذت عينة أخرى مستقلة عن الأولى حجمها $n = 10$ ، وتباينها $S_2^2 = 120$ ، من توزيع طبيعي $N\left(\mu_2 , \sigma_2^2\right)$.

أوجد 90 % فترة ثقة للنسبة $\dfrac{\sigma_2^2}{\sigma_1^2}$.

الحل : لدينا : $1-\alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow \dfrac{\alpha}{2} = 0.05 \Rightarrow 1-\dfrac{\alpha}{2} = 0.95$

أنّ القيمتين المعياريتين هما :

$$f_{1-\frac{\alpha}{2}}(m-1, n-1) = f_{0.95}(5,9) = 3.48$$

$$f_{\frac{\alpha}{2}}(m-1, n-1) = f_{0.05}(5,9) = 0.21$$

أنّ فترة الثقة للنسبة المطلوبة تعطى بالشكل :

$$C.I. = \left[\frac{S_2^2}{S_1^2} f_{\frac{\alpha}{2}}(m-1 , n-1) \;\; ; \;\; \frac{S_2^2}{S_1^2} f_{1-\frac{\alpha}{2}}(m-1 , n-1) \right]$$

$$= \left[\frac{120}{80}(0.21) \;\; ; \;\; \frac{120}{80}(3.48) \right] = [0.315 \;\; ; \;\; 5.22]$$

مبرهنة (10) : لتكن $X_1 , X_2 , \ldots , X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الأسي $f(x,\vartheta) = \vartheta e^{-\vartheta x} \;\; , \;\; x > 0$. أنّ $100(1-\alpha)\%$ فترة ثقة للوسيط ϑ تعطى بالفترة $[L_1 , L_2]$ ، حيث :

$$L_1 = \frac{\chi^2_{\frac{\alpha}{2}}(2n)}{2n\,\overline{X}} \quad ; \quad L_2 = \frac{\chi^2_{1-\frac{\alpha}{2}}(2n)}{2n\,\overline{X}}$$

أنّ برهان هذه المبرهنة ليس من منهاج هذا الكتاب ، حيث أنّ لهذه النظرية تطبيق مهم جداً في مجال الوثوقية .

الفصل الرّابع

اختبار الفرضيات الإحصائية

Test of statistical hypothesis

أولاً- مقدمة: أنّ البحث المهم من أبحاث الاستقراء الإحصائي ، والذي سنعالجه في هذا الفصل هو اختبار الفرضيات الإحصائية ، وذلك لما لهذا الموضوع من أهمية خاصة في اتخاذ القرارات الإحصائية بقبول أو رفض تلك الفرضيات . ونحن نعلم أنّ للإحصاء أهمية بالغة في تحديد سير العمل واتخاذ القرارات الحاسمة والسريعة في مواقف تتحكم فيها المصادفة ويتطلب اتخاذ القرار شيئاً من العقلانية ، وتقديراً كمياً للمخاطرة .

ثانياً- الفرضية الإحصائية: هي كل إفادة أو ادعاء أو مقولة تتعلق بالمجتمع الإحصائي تحتمل الصحة أو الخطأ . أنّ صحة الفرضية الإحصائية أو خطأها لا يمكن معرفته بدقة تامة إلا إذا تفحصنا المجتمع الإحصائي بكامله ، وهذا غير ممكن في معظم الأحيان ، لذلك نختار عينة عشوائية من المجتمع الإحصائي المدروس ونتفحص هذه العينة العشوائية بشكل جيد وذلك بغية استخدام المعلومات التي تحويها في اتخاذ القرارات المناسبة فيما إذا كانت الفرضية الإحصائية صحيحة أو خاطئة ، لكن هذا القرار الذي نتخذه لا يكون سليماً دوماً ، فقد تخدعنا النتائج التي نحصل عليها من تلك العينة ، فنرفض فرضية صحيحة ، أو نقبل فرضية خاطئة لذلك فإنّ قبولنا لفرضية إحصائية ما لا يعني أنّها صحيحة تماماً ، كما أنّ رفضنا لها لا يعني أنّها خاطئة تماماً ، ولهذا السبب فإنّ الإحصائي يختبر دوماً الفرضية التي يأمل أنّ يرفضها ، فمثلاً لإثبات أنّ طريقة ما أفضل من طريقة أخرى في استخلاص غاز النشادر، نختبر الفرضية الآتية: لا فرق بين الطريقتين.

سوف نرمز للفرضية التي نريد اختبارها بالفرضية الابتدائية أو الفرضية الصفرية ونرمز لها بالرمز H_0 ، وأنّ رفضنا للفرضية الابتدائية H_0 يقودنا إلى قبول فرضية أخرى نسميها بالفرضية البديلة ونرمز لها بالرمز H_1 .

فإذا أردنا اختبار فرضية تتعلق بقيم وسيط مجتمع إحصائي ، لا بد لنا من تحديد الفرضية الابتدائية ومن ثم حساب إحصاء الاختبار ، وبعد ذلك نحدد منطقة القبول أو منطقة الرفض، وأخيراً الفرضية البديلة .

ونظراً لاختلاف قيم إحصاء الاختبار Z باختلاف العينة العشوائية المختارة يبدو من المعقول أنّ نقبل الفرضية الابتدائية إذا كانت قيم إحصاء الاختبار Z قريبة نسبياً من القيمة المتوقعة لها في حالة صحة هذه الفرضية ، أما

إذا كانت نواتج الاختبار لا تتوافق مع هذه القيمة ، نقول عندئذٍ أنّ الفرق ذو أهمية ونرفض الفرضية الابتدائية H_0 . وعند قبولنا الفرضية الابتدائية أو رفضنا لها يمكن أنّ نرتكب خطأ وهذا الخطأ على نوعين :

1 – الخطأ من النوع الأول: وهو الخطأ الناجم من رفض الفرضية الابتدائية H_0 على الرغم من أنّها صحيحة ويدعى احتمال الوقوع في الخطأ من النوع الأول بمستوى الدلالة أو مستوى الأهمية ونرمز له بالرمز α .

2 – الخطأ من النوع الثاني: وهو الخطأ الناجم عن قبول الفرضية الابتدائية H_0 على الرغم من أنّها خاطئة ، نرمز لاحتمال الوقوع في الخطأ من النوع الثاني بالرمز β . بينما يدعى $(1-\beta)$ بقوة اختبار الفرضية الابتدائية H_0 مقابل الفرضية البديلة H_1 .

أنّ الشكل الآتي يوضح الحالات الممكنة للفرضية الابتدائية ونوعي القرار الممكنين (قبول أو رفض) بالإضافة إلى نوعي الخطأ اللذين يمكن ارتكابهما (الخطأ من النوع الأول أو الخطأ من النوع الثاني) .

القرار / الفرضية	قبول الفرضية	رفض الفرضية
صحيحة	قرار سليم وصحيح	خطأ من النوع الأول
خاطئة	خطا من النوع الثاني	قرار سليم وصحيح

ثالثاً– خطوات اختبار الفرضية:

إذا أردنا اختبار فرضية إحصائية تتعلق بوسيط مجتمع إحصائي أو أكثر، لابد من القيام بالخطوات الآتية :

1– تحديد الفرضية الابتدائية H_0 والفرضية البديلة H_1 التي سنقبلها في حال رفضنا للفرضية H_0 .

2– اختيار عينة عشوائية من المجتمع الإحصائي ثم إيجاد القيم المميزة لهذه العينة العشوائية مثل متوسط العينة ، تباين العينة ، ... .

3– إيجاد إحصاء الاختبار Z وذلك بالاعتماد على المتغيرات المعيارية التي استخدمناها في التقدير النقطي والمجالي لوسطاء المجتمع الإحصائي في الفصل السابق وحيث يحسب إحصاء الاختبار Z من العلاقة : $Z = \dfrac{\hat{\theta} - \theta_0}{\hat{\sigma}_{(\hat{\theta})}}$ حيث $\hat{\theta}$ هو التقدير النقطي للوسيط المجهول θ ويحسب من المعلومات المتوفرة حول العينة العشوائية المختارة من المجتمع الإحصائي مثل متوسط العينة ، تباين العينة . $\hat{\sigma}_{(\hat{\theta})}$ هو مقدِّر الانحراف المعياري للمقدِّر $\hat{\theta}$.

θ_0 هي قيمة θ المأخوذة من الفرضية الابتدائية وهي معلومة دوماً.

4– تحديد مستوى الأهمية α ، وأنشاء منطقة الرفض التي مساحتها تساوي α وهي المنطقة التي تنتمي إليها إحصاءات العينة المأخوذة من المجتمع الإحصائي والتي تختلف عما تحدده الفرضية الابتدائية.

5– اتخاذ القرار السليم والصحيح في قبول أو رفض الفرضية الابتدائية H_0 ، فإذا وقع Z في منطقة القبول ، فإنّنا نقبل الفرضية الابتدائية H_0 وإلا رفضناها .

رابعاً- العوامل المؤثرة في قوة الاختبار:

تتأثر قوة الاختبار بعدة عوامل هي :

1- حجم العينة: تزداد قوة الاختبار لقيمة معينة للوسيط المراد اختباره بازدياد حجم العينة ، وهنا يجب الانتباه إلى أنّ العينات كبيرة الحجم تجعل أي فرق بسيط بين الإحصاء والوسيط المقابل له فرقاً ذا دلالة إحصائية وربما تركنا ذلك في حيرة من الأمر ، إذ ربما لا يكون لهذا الفرق أهمية عملية في حين أنّ الاختبار الإحصائي يشير إلى أنّه يوجد فرق جوهري .

2- مستوى المعنوية: تزداد قوة الاختبار بقيمة معينة للوسيط المراد اختباره بازدياد قيمة مستوى الدلالة . فلو مثلنا هذا بيانياً لتبين لنا أثر مستوى الدلالة على قوة الاختبار ، فزيادة قيمة α تعني انزياح الحد الحرج نحو اليسار ، وينتج عن ذلك نقصان قيمة β وبالتالي زيادة $(1-\beta)$ أي زيادة قوة الاختبار .

3- علاقة القيمة الحقيقية للوسيط بقيمته في الفرضية الابتدائية : بفرض أنّ الاختبار باتجاهين ، فإنّ قوة الاختبار تزداد كلما ابتعدت القيمة الحقيقية للوسيط عن القيمة المفروضة .

نلاحظ أنّه إذا رسمنا الشكل المناسب فإنّنا نجد بوضوح علاقة قوة الاختبار بابتعاد متوسط المجتمع الطبيعي μ عن القيمة المفروضة بالفرضية الابتدائية لاختبار باتجاه واحد أو باتجاهين .

وكما هو واضح فإنّ قوة الاختبار تزداد كثيراً بابتعاد القيمة الحقيقية عن القيمة المفروضة للجهتين ، وتكون قوة الاختبار في نهايتها الصغرى عندما تكون القيمة الحقيقية مساوية تماماً للقيمة المفروضة.

4- كون الاختبار باتجاه واحد أو باتجاهين: يكون الاختبار باتجاه واحد أقوى من الاختبار باتجاهين إذا كانت القيمة الحقيقية للوسيط في نفس الجهة التي تفترضها الفرضية البديلة.

خامساً- اختبار الوسطاء الإحصائية:

سوف ندرس في هذه الفترة اختبار الوسطاء الإحصائية وسوف نميز بين أنّواع الاختبارات وندرس فقط الاختبارات المتعلقة بوسطاء المجتمع الطبيعي والمجتمع الحداني من الشكل $b\ (1\ ,\ P)$ ، لذلك نبدأ بتعريف أنّواع الاختبارات

تعريف: الاختبار وحيد الذيل (أو باتجاه واحد): وهو اختبار الفرضية الابتدائية: $H_0 : \theta = \theta_0$ ضد الفرضية البديلة: $H_1 : \theta > \theta_0$ أو هو اختبار الفرضية الابتدائية: $H_0 : \theta = \theta_0$ ضد الفرضية البديلة: $H_1 : \theta < \theta_0$.

تعريف: الاختبار ثنائي الذيل (أو باتجاهين): وهو اختبار الفرضية الابتدائية : $H_0 : \theta = \theta_0$ ضد الفرضية البديلة $H_1 : \theta \neq \theta_0$.

أولاً – اختبار متوسط مجتمع طبيعي تباينه σ^2 معلوم:

لقد درسنا في الفصل السابق كيفية إيجاد فترات الثقة لمتوسط مجتمع طبيعي وللفرق بين متوسطي مجتمعين طبيعيين ، وفي هذه الفقرة سوف نقوم بإجراء اختبارات حول هذه الوسطاء الإحصائية .

مبرهنة (1): لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu, \sigma^2)$ ، حيث σ^2 معلومة وبفرض أنّ μ_o معلومة وأنّ α مستوى الدلالة .

آ – لاختبار الفرضية: $H_o : \mu = \mu_o$ ضد الفرضية: $H_1 : \mu > \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان:

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_o\right)}{\sigma} > z_{1-\alpha}$$

ب – لاختبار الفرضية : $H_o : \mu = \mu_o$ ضد الفرضية : $H_1 : \mu < \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان:

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_o\right)}{\sigma} < -z_{1-\alpha}$$

ج – لاختبار الفرضية : $H_o : \mu = \mu_o$ ضد الفرضية : $H_1 : \mu \neq \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان:

$$\left. \begin{array}{l} \leq -z_{1-\frac{\alpha}{2}} \\ \text{أو} \\ \geq +z_{1-\frac{\alpha}{2}} \end{array} \right\} Z = \frac{\sqrt{n}\left(\overline{X} - \mu_o\right)}{\sigma} \quad \text{أي إذا كان:} \quad |Z| > z_{1-\frac{\alpha}{2}}$$

مثال: لتكن $X_1, X_2, \ldots, X_{25}$ عينة عشوائية مأخوذة من مجتمع طبيعي $N(\mu, 16)$ ، بحيث أنّ $\sum_{i=1}^{25} X_i = 250$.

اختبر الفرضية: $H_o : \mu = 9$ ضد الفرضية : $H_1 : \mu \neq 9$ عند مستوى الأهمية $\alpha = 0.05$.

الحل : لدينا $\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, ولدينا إحصاء الاختبار Z معرّف بالشكل :

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_o\right)}{\sigma} = \frac{5(10-9)}{4} = \frac{5}{4} = 1.25$$

نلاحظ أنّ $Z = 1.5 < 1.96 = z_{0.975}$ وبالتالي نقبل الفرضية H_0 عند مستوى الأهمية $\alpha = 0.05$.

مثال: استخدم بيانات المثال السابق في اختبار الفرضية: $H_o : \mu = 8$ ضد الفرضية : $H_1 : \mu > 8$.

الحل: لدينا $\alpha = 0.05 \Rightarrow 1-\alpha = 0.95 \Rightarrow z_{1-\alpha} = z_{0.95} = 1.64$, ولدينا إحصاء الاختبار Z معرّف بالشكل:

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_o\right)}{\sigma} = \frac{5(10-8)}{4} = \frac{10}{4} = 2.5$$

نلاحظ أنّ $Z = 2.5 > 1.64 = z_{0.95}$, وبالتالي نقبل الفرضية H_0 عند مستوى الأهمية $\alpha = 0.05$.

مثال: تدل الدراسات أنّ وزن الطفل عند الولادة يخضع للتوزيع الطبيعي متوسطه 3.2 كغ وانحرافه المعياري 0.6 كغ . وعند معاينة 49 طفلاً خضعت أمهاتهم لنظام غذائي جديد تبين أنّ متوسط الوزن لهذه العينة 3.5 كغ . هل تستنتج أنّ النظام الغذائي الجديد قد ساهم في تحسين متوسط وزن الطفل بمستوى $\alpha = 0.025$ من الأهمية .

الحل: لنكوّن الفرضية الابتدائية: $H_0 : \mu = 3.2$, ومن ثمَّ فإنّ الفرضية البديلة : $H_1 : \mu > 3.2$.

لدينا $\alpha = 0.025 \Leftarrow 1-\alpha = 0.975 \Leftarrow z_{0.975} = 1.96$ ولدينا: $Z = \dfrac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{\sigma} = 3.5$

نلاحظ أنّ : $Z = 3.5 > 1.96 = z_{0.975}$, من ثمَّ فإنّنا نرفض الفرضية H_o وهذا يعني أنّ النظام الغذائي الجديد قد حسَّن في وزن الطفل عند ولادته .

ثانياً – اختبار متوسط مجتمع طبيعي تباينه σ^2 مجهول :

مبرهنة (2) : لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية مأخوذة من المجتمع الطبيعي $N\left(\mu,\sigma^2\right)$ ، حيث σ^2 مجهولة ، وبفرض أنّ μ_o معلومة , S^2 هو تباين العينة وأنّ α مستوى الدلالة .

آ – لاختبار الفرضية: $H_o : \mu = \mu_o$ ضد الفرضية: $H_1 : \mu > \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$T = \frac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{S} > t_{1-\alpha}(n-1)$$

ب – لاختبار الفرضية: $H_o : \mu = \mu_o$ ضد الفرضية : $H_1 : \mu < \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$T = \frac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{S} < -t_{1-\alpha}(n-1)$$

ج – لاختبار الفرضية : $H_o : \mu = \mu_o$ ضد الفرضية : $H_1 : \mu \neq \mu_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$T = \frac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{S} \begin{cases} \leq -t_{1-\frac{\alpha}{2}}(n-1) \\ \text{أو} \\ \geq +t_{1-\frac{\alpha}{2}}(n-1) \end{cases}$$

أي إذا كان: $|T| \geq t_{1-\frac{\alpha}{2}}(n-1)$

مثال: لتكن $X_1, X_2, \ldots, X_{16}$ عينة عشوائية مأخوذة من مجتمع طبيعي بحيث أنّ $\overline{X} = 80$, $S^2 = 25$.

والمطلوب: اختبر الفرضية: $H_o : \mu = 75$ ضد الفرضية : $H_1 : \mu < 75$ عند مستوى الدلالة $\alpha = 0.10$.

الحل : لدينا $\alpha = 0.10 \Rightarrow 1-\alpha = 0.90 \Rightarrow t_{1-\alpha}(n-1) = t_{0.90}(15) = 1.341$

لنوجد إحصاء الاختبار T بالشكل : $T = \dfrac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{S} = \dfrac{4(80-75)}{5} = 4$

نلاحظ أنّ: $T = 4 > 1.341 = t_{1-\alpha}(n-1)$ ومن ثمَّ فإنّنا نرفض الفرضية H_0 عند مستوى الدلالة $\alpha = 0.10$.

مثال: آلة لقطع رقائق معدنية أنْتجت في الماضي قطعاً سمكها 0.5 ميليمتر ، ولتقدير ما إذا كانت الآلة تعمل بشكل مرضٍ ، سحبت عينة عشوائية حجمها n = 9 ووجد أنّ معدل سمكها 0.53 ميليمتر وانحرافها المعياري 0.03 ميليمتر . وبفرض أنّ سماكة القطع المنتجة تخضع للتوزيع الطبيعي والمطلوب:

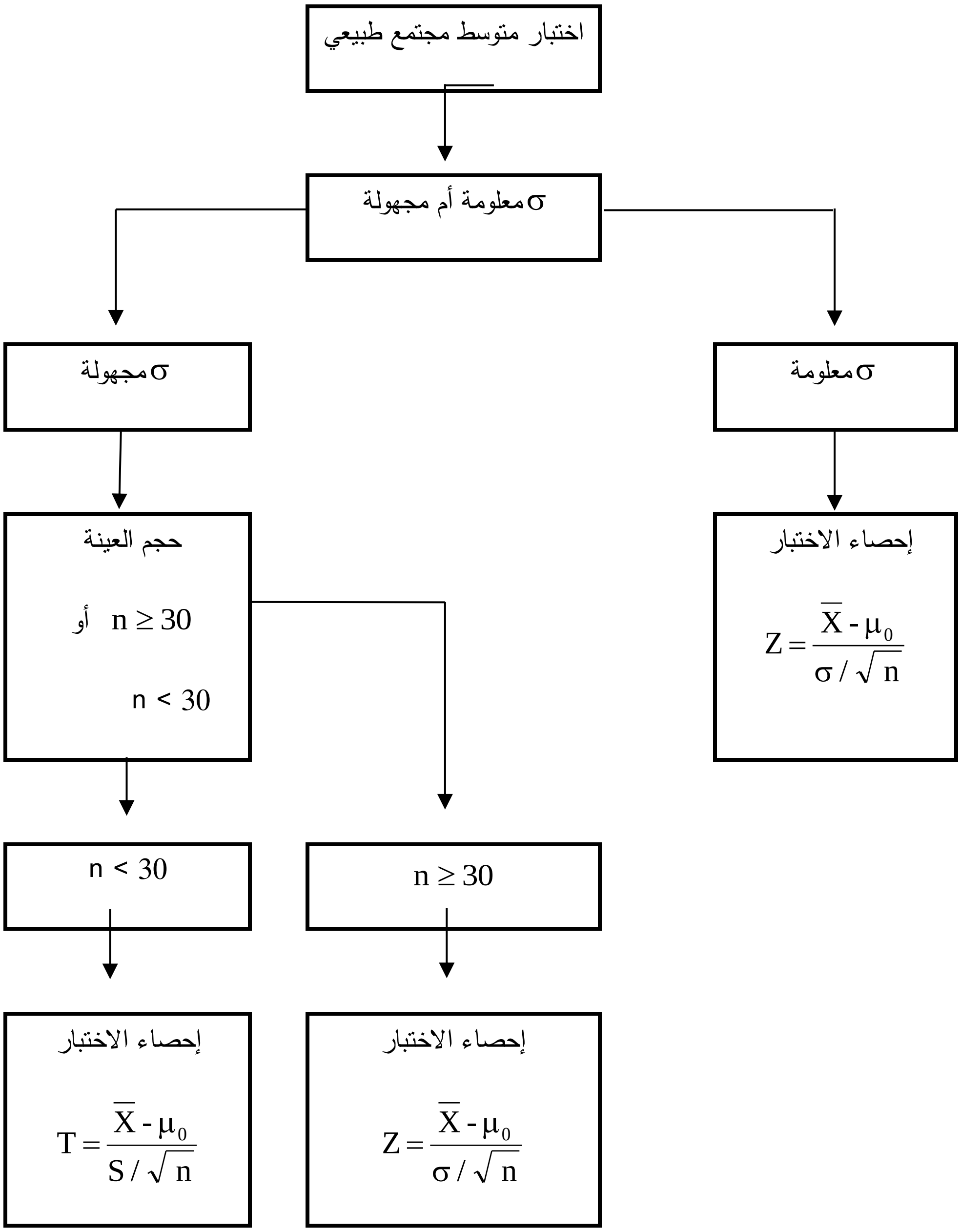
اختبر الفرضية: $H_o : \mu = 0.5$ ضد الفرضية : $H_1 : \mu > 0.50$ عند مستوى الدلالة $\alpha = 0.05$.

الحل: لدينا $\alpha = 0.05$ ومنه نجد $1-\alpha = 0.95$ ، ولدينا $n-1 = 8$ وبالتالي $t_{1-\alpha}(n-1) = t_{0.95}(8) = 1.86$

ولدينا: $3 = \frac{3(0.53-0.50)}{0.03} = \frac{\sqrt{n}\,(\bar{X}-\mu_o)}{S} = T$, بالمقارنة نجد أنّ: $T = 3 > 1.86 = t_{0.95}(8)$.

وبالتالي فإنّنا نرفض الفرضية H_0 عند مستوى الأهمية $\alpha = 0.05$.

سوف نوضح فيما يلي مخططاً في اختبار متوسط مجتمع طبيعي:



مثال: تبين لدى قياس معدل البولة في الدم لـ 25 شخصاً مختارين من مجتمع طبيعي أنّ متوسط هذه العينة هو 0.29 غ/ل وأنّ انحرافها المعياري $S = 0.05$ غ/ل والمطلوب:

اختبر الفرضية: $H_o : \mu = 0.3$ ضد الفرضية : $H_1 : \mu < 0.03$ عند مستوى الأهمية $\alpha = 0.05$.

الحل: لدينا $t_{1-\alpha}(n-1) = t_{0.95}(24) = 1.711$ وأنّ : $-1 = \frac{5(0.29-0.3)}{0.05} = \frac{\sqrt{n}\,(\bar{X}-\mu_o)}{S} = T$.

نلاحظ أنّ : $T = -1 < -1.711 = -t_{0.95}(24)$. ليست محققة وبالتالي فإنّنا نقبل الفرضية H_o .

مثال: صممت آلة لإنتاج الصفائح المعدنية بحيث تنتج صفائح ثخانتها تخضع للتوزيع الطبيعي متوسطه 3 ملم . ولاختبار جودة هذه الآلة تم معاينة 9 صفائح من انتاجها فتبين أنّ متوسط الثخانة للعينة $\overline{X} = 2.8$ ملم وانحرافها $S = 0.55$ ملم. فهل ما زالت تعمل هذه الآلة بالمواصفات التي صممت من أجلها بمستوى الأهمية $\alpha = 0.05$.

الحل: لنشكل الفرضية الابتدائية بالشكل : $H_o : \mu = 3$

وبالتالي فإنّ الفرضية البديلة هي : $H_1 : \mu \neq 3$

لدينا: $t_{0.975}(8) = 2.306$ ولدينا $T = \frac{\sqrt{n}\left(\overline{X}-\mu\right)}{S} = \frac{3(2.8-3)}{0.55} = -1.09$.

نلاحظ أنّ : $-2.306 < T = -1.09 < 2.306$

وبالتالي فإنّنا نقبل الفرضية H_o عند مستوى الأهمية $\alpha = 0.05$ أي أنّ الآلة مازالت تعمل بالمواصفات المطلوبة .

ثالثاً – اختبار الفرق بين متوسطي مجتمعين طبيعيين تبايناهما معلومان:

مبرهنة (3): لتكن $X_1, X_2, ..., X_m$ عينة عشوائية مأخوذة من المجتمع الطبيعي $N\left(\mu_1, \sigma_1^2\right)$ ، ولتكن $Y_1, Y_2, ..., Y_n$ عينة عشوائية مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N\left(\mu_2, \sigma_2^2\right)$ ، وبفرض أنّ α هي مستوى الدلالة :

آ – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 > \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} > z_{1-\alpha}$$

ب – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 < \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} < -z_{1-\alpha}$$

ج – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 \neq \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \begin{cases} \leq -z_{1-\frac{\alpha}{2}} \\ \text{أو} \\ \geq +z_{1-\frac{\alpha}{2}} \end{cases}$$

أي إذا كان: $|Z| \geq z_{1-\frac{\alpha}{2}}$

مثال: إذا أخذنا عينتين مستقلتين حجمهما $n = 50,\ m = 75$ من مجتمعين طبيعيين $N(\mu_1, 64)$, $N(\mu_2, 36)$ وكان متوسطا هاتين العينتين 82 و 76 على الترتيب فالمطلوب :

اختبر الفرضية $H_o : \mu_1 = \mu_2$ ضد الفرضية : $H_1 : \mu_1 < \mu_2$ عند مستوى الأهمية $\alpha = 0.05$.

الحل: أنّ منطقة الرفض هي: $Z \leq - z_{1-\alpha}$ ولدينا $z_{1-\alpha} = z_{0.95} = 1.64$.

لنوجد الآن طرفي هذه المتباينة : $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} = \dfrac{82 - 76}{\sqrt{\dfrac{64}{75} + \dfrac{36}{50}}} = 4.78$

نلاحظ أنّ المتباينة ليست محققة وذلك لأنّ المتباينة : $Z = 4.78 \leq - 1.64 = -z_{1-\alpha}$.

ليست صحيحة ، ومن ثمَّ نقبل الفرضية H_0 عند مستوى الأهمية $\alpha = 0.05$.

مثال: لنفترض أنّنا نرغب في مقارنة نوعين من أنّظمة التغذية لمجموعتين مختلفتين من الأطفال حديثي الولادة لتبيأنّ أي من النظامين يعطي زيادة أكبر في وزن الطفل فإذا كان لنظامي التغذية التوزيعين الطبيعيين $N(\mu_1 , 300)$ و $N(\mu_2 , 400)$ على الترتيب وطبقنا نظام التغذية الأول على 50 طفلاً فكان متوسط الزيادة خلال أسبوع 340 غراماً ثم طبقنا نظام التغذية الآخر على 40 طفلاً فكان متوسط الزيادة خلال أسبوع 325 غراماً.

فهل تستنتج أنّ نظام التغذية الأول يسبب زيادة أكبر في وزن الأطفال عند مستوى الأهمية $\alpha = 0.10$.

الحل : لنشكل الفرضية الابتدائية بالشكل : $H_o : \mu_1 = \mu_2$

وبالتالي فإنّ الفرضية البديلة هي : $H_1 : \mu_1 > \mu_2$

أنّ منطقة الرفض هي: $Z \geq z_{1-\alpha}$

لنوجد الآن طرفي هذه المتباينة : $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} = \dfrac{340 - 325}{\sqrt{\dfrac{300}{50} + \dfrac{400}{40}}} = \dfrac{15}{4} = 3.75$

ولدينا: $z_{1-\alpha} = z_{0.90} = 1.28$. نلاحظ أنّ: $Z = 3.75 > 1.28 = z_{1-\alpha}$. وبالتالي نرفض الفرضية H_0 وهذا يعني أنّ نظام التغذية الأول يسبب زيادة أكبر في وزن الأطفال بمستوى $\alpha = 0.10$ من الأهمية .

رابعاً – اختبار الفرق بين متوسطي مجتمعين طبيعيين تبايناهما مجهولان :

مبرهنة (4): لتكن $X_1 , X_2 ,..., X_m$ عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu_1 , \sigma_1^2)$ ، ولتكن $Y_1 , Y_2 ,..., Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N(\mu_2 , \sigma_2^2)$ ، حيث $\sigma_1^2 = \sigma_2^2$ مجهولان، وبفرض أنّ S_P^2 هو التباين المكافئ والمعرّف بالعلاقة :

$S_P^2 = \dfrac{(m-1)\, S_1^2 + (n-1)\, S_2^2}{m+n-2}$, حيث S_1^2 , S_2^2 هما تباينا العينتين على الترتيب . α هي مستوى الدلالة .

آ – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 > \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$T = \frac{\overline{X} - \overline{Y}}{S_P\sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{1-\alpha}(m+n-2)$$

ب – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 < \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$T = \frac{\overline{X} - \overline{Y}}{S_P\sqrt{\frac{1}{m} + \frac{1}{n}}} < -t_{1-\alpha}(m+n-2)$$

ج – لاختبار الفرضية: $H_o : \mu_1 = \mu_2$ ضد الفرضية: $H_1 : \mu_1 \neq \mu_2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \begin{cases} \leq -t_{1-\frac{\alpha}{2}}(m+n-2) \\ \text{أو} \\ \geq +t_{1-\frac{\alpha}{2}}(m+n-2) \end{cases}$$

، أي إذا كان: $|T| \geq t_{1-\frac{\alpha}{2}}(m+n-2)$

مثال:

في دراسة حول مرض الخناق عند الجرذان ، أخذت عينتان عشوائيتان حجم كل منهما 9 جرذان . أعطيت المجموعة الأولى جرعة من دواء البليسبوا (اسمه Placebo) وهو دواء وهمي لا يضرّ ولا ينفع. وأعطيت المجموعة الثانية دواءً تجريبياً اسمه FL113 . وبعد ذلك أجريت الفحوصات اللازمة على هذه الجرذان مع مراعاة أنّ وقت التحسن لكل جرذ قد حدد . وقد اعتقد أنّ الدواء التجريبي سوف يقلل زمن المرض ، وبفرض أنّ مجتمعي الجرذان هما $N(\mu_1, \sigma_1^2)$ و $N(\mu_2, \sigma_2^2)$ وهما مستقلان وبعد التجربة حصلنا على المعلومات الآتية :

$\overline{X}_2 = 283$ ثانية ، $\overline{X}_1 = 329$ ثاني ، $m = n = 9$ ، $S_2 = 43$ ثانية ، $S_1 = 45$ ثانية والمطلوب :

اختبر الفرضية الابتدائية: $H_o : \mu_1 = \mu_2$ ضد الفرضية البديلة: $H_1 : \mu_1 > \mu_2$ عند مستوى الأهمية $\alpha = 0.01$.

الحل: أنّ منطقة الرفض هي $T \geq t_{1-\alpha}(m+n-2)$

لنوجد كلاً من طرفي هذه المتباينة : لدينا إحصاء الاختبار

$$T = \frac{\overline{X} - \overline{Y}}{S_P\sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{329 - 283}{\sqrt{\frac{88}{9}}} = 14.71$$

حيث أنّ التباين المكافئ هو:

$$S_P^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} = \frac{8(45)^2 + 8(43)^2}{16} = 44$$

ولدينا $t_{1-\alpha}(m+n-2) = t_{0.90}(16) = 2.583$.

نلاحظ أنّ: $T = 14.71 > 2.583 = t_{0.91}(16)$ وبالتالي نرفض الفرضية H_o عند مستوى الأهمية $\alpha = 0.01$.

خامساً – اختبار النسبة P في التوزيع الحداني (b (x ; 1 , P :

كثيراً ما يصادفنا في المجتمعات الحدانية ، في مجالات متعددة ، معرفة نسبة نجاح عملية جراحية ، أو معرفة

نسبة الأشخاص المقترعين لصالح المرشح A ، أو معرفة النسبة المئوية لفعالية لقاح جديد ضد مرض معين . فقد تكون هذه النسبة مقبولة أو غير مقبولة ، لذلك علينا اتخاذ القرار المناسب في قبول أو رفض هذه النسبة . لنوجد الآن اختبار الفرضية المناسب .

مبرهنة (5): لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من المجتمع الحداني $b(x\,;1\,,P)$ وبفرض أنّ n كبيرة وأنّ P_0 عدد موجب ، α هي مستوى الأهمية.

آ – لاختبار الفرضية: $H_o : P = P_o$ ضد الفرضية : $H_1 : P \neq P_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P} - P_0}{\sqrt{\dfrac{P_o(1-P_o)}{n}}} \begin{cases} \leq -z_{1-\frac{\alpha}{2}} \\ \text{أو} \\ \geq +z_{1-\frac{\alpha}{2}} \end{cases}$$

ب – لاختبار الفرضية : $H_o : P = P_o$ ضد الفرضية : $H_1 : P > P_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P} - P_0}{\sqrt{\dfrac{P_o(1-P_o)}{n}}} \geq z_{1-\alpha}$$

ج – لاختبار الفرضية : $H_o : P = P_o$ ضد الفرضية : $H_1 : P < P_o$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P} - P_0}{\sqrt{\dfrac{P_o(1-P_o)}{n}}} \leq -z_{1-\alpha}$$

مثال: أرادت شركة طبية أنّ تجري اختباراً لمعرفة مدى فعالية دواء جديد لمرض معين . أخذت عينة عشوائية حجمها n = 1600 من المصابين بهذا المرض. فإذا أعطي هذا الدواء لهؤلاء المصابين وتبين بعد ذلك أنّ الذين تمَّ شفاؤهم بتأثير هذا الدواء 1200 مريض. فهل نقبل عند مستوى الأهمية $\alpha = 0.10$ بادعاء الشركة المنتجة بأنّ % 80 من الذين يعالجون بهذا الدواء يتم شفاؤهم .

الحل : نشكل الفرضية الابتدائية $H_o : P = 0.80$

ثم نشكل الفرضية البديلة $H_1 : P > 0.80$

لنوجد إحصاء الاختبار Z :

$$Z = \frac{\hat{P} - P_0}{\sqrt{\dfrac{P_o(1-P_o)}{n}}} = \frac{\dfrac{1200}{100} - 0.80}{\sqrt{\dfrac{(0.8)(0.2)}{1600}}} = \frac{0.75 - 0.80}{\sqrt{0.0001}} = \frac{-0.5}{0.01} = -5$$

ولدينا $z_{1-\alpha} = z_{0.90} = 1.285$. بالمقارنة نجد أنّ: $Z = -5 > 1.285$ وبالتالي نقبل الفرضية H_o .

سادساً – اختبار الفرق $(P_1 - P_2)$ بين نسبتي توزيعين حدانيين :

تبرز أهمية المقارنة بين نسبتي مجتمعين حدانيين $b(x;1,P_1)$, $b(x;1,P_2)$ عندما نريد اختبار فرضية تساوي نسبتين ، فمثلاً يمكننا إجراء اختبار ما إذا كانت نسبة مدرسي مادة الرياضيات تفوق النسبة المطابقة في الفيزياء . وكمثال آخر ، يمكن لمدخن أنّ يقرر ما إذا كان عليه أنّ يقلع عن التدخين ، إذا اقتنع أنّ نسبة المصابين بورم رئوي تفوق نسبة غير المدخنين المصابين بهذا المرض .

مبرهنة (6): لتكن $X_1, X_2, \dots, X_m$ عينة عشوائية مأخوذة من المجتمع الحداني $b(1\ ,\ P_1)$ ولتكن $Y_1, Y_2, \dots, Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من مجتمع حداني آخر $b(1\ ,\ P_2)$ ، وبفرض أنّ m و n كبيرتين , α هي مستوى الأهمية و $\hat{P} = \dfrac{m\hat{P}_1 + n\hat{P}_2}{m+n}$.

آ – لاختبار الفرضية: $H_o : P_1 = P_2$ ضد الفرضية : $H_1 : P_1 \neq P_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} \begin{cases} \leq -z_{1-\frac{\alpha}{2}} \\ \text{أو} \\ \geq +z_{1-\frac{\alpha}{2}} \end{cases}$$

ب – لاختبار الفرضية: $H_o : P_1 = P_2$ ضد الفرضية : $H_1 : P_1 > P_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} > z_{1-\alpha}$$

ج – لاختبار الفرضية: $H_o : P_1 = P_1$ ضد الفرضية : $H_1 : P_1 < P_2$ فإنّنا نرفض الفرضية H_o إذا كان :

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} < -z_{1-\alpha}$$

مثال: جرى استفتاء شعبي في مدينة درعا ، والريف التابع لها حول إمكانية أنشاء ملعب رياضي في أدنى المدينة فإذا صوّت لصالح المشروع 120 شخصاً من أصل 200 من سكان المدينة و 240 شخصاً من أصل 500 من سكان الريف لصالح هذا الاستفتاء . فهل نستنتج أنّ نسبة المصوتين لصالح هذا الاستفتاء من سكان المدينة تفوق مثيلتها من سكان الريف بمستوى $\alpha = 0.10$ من الأهمية .

الحل : نشكل الفرضية الابتدائية: $H_o : P_1 = P_2$

ثم نشكل الفرضية البديلة : $H_1 : P_1 > P_2$

وبالتالي فإنّ منطقة الرفض هي : $Z \geq z_{1-\alpha}$

لنوجد إحصاء الاختبار Z : $-3 = $
$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\frac{1}{m}+\frac{1}{n}\right)}} = \frac{0.6 - 0.48}{\sqrt{(0.51)(0.49)\left(\frac{1}{200}+\frac{1}{500}\right)}} = -3$$

حيث: $\hat{P}_1 = \frac{120}{200} = 0.61 \quad , \quad \hat{P}_2 = \frac{240}{500} = 0.48 \quad , \hat{P} = \frac{m\hat{P}_1 + n\hat{P}_2}{m+n} = \frac{120+240}{200+500} = 0.51$

ثم نوجد القيمة المعيارية $z_{1-\alpha}$: $z_{1-\alpha} = z_{1-0.10} = z_{0.90} = 1.285$

بالمقارنة نجد أنّ : $Z = 3 > 1.285 = z_{0.90}$, مما يبين أنّنا سوف نرفض الفرضية H_0 عند مستوى الأهمية $\alpha = 0.10$ ونقول أنّ نسبة المقترعين لصالح الاستفتاء من سكان المدينة تفوق مثيلتها من سكان الريف .

سابعاً – اختبار تباين مجتمع طبيعي :

نحن نعلم أنّ المتغير العشوائي $\frac{(n-1)S^2}{\sigma^2}$ يخضع لتوزيع كاي مربع درجة حريته (n – 1) ، حيث $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ ، وأنّ $100(1-\alpha)\%$ فترة الثقة لتباين مجتمع طبيعي هي :

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \; ; \; \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$$

مبرهنة (7): لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية مأخوذة من مجتمع طبيعي $N\left(\mu, \sigma^2\right)$. حيث σ_0^2 معلوم .

آ – لاختبار الفرضية: $H_0 : \sigma^2 = \sigma_0^2$ ضد الفرضية: $H_1 : \sigma^2 \neq \sigma_0^2$ فإنّنا نرفض الفرضية H_0 إذا كان:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \begin{cases} \leq \chi^2_{\frac{\alpha}{2}}(n-1) \\ \text{أو} \\ \geq \chi^2_{1-\frac{\alpha}{2}}(n-1) \end{cases}$$

ب – لاختبار الفرضية: $H_0 : \sigma^2 = \sigma_0^2$ ضد الفرضية: $H_1 : \sigma^2 > \sigma_0^2$ فإنّنا نرفض الفرضية H_0 إذا كان:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{1-\alpha}(n-1)$$

ج – لاختبار الفرضية: $H_0 : \sigma^2 = \sigma_0^2$ ضد الفرضية: $H_1 : \sigma^2 < \sigma_0^2$ فإنّنا نرفض الفرضية H_0 إذا كان:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha}(n-1)$$

مثال: لدى مجرب القناعة بأنّ لجهاز القياس الذي يستخدمه تغيراً مقيساً بالانحراف المعياري $\sigma = 2$. وقد سجل خلال تجربة القياسات 4.1 , 5.2 , 10.2 فهل تتناقص هذه القياسات مع قناعته عن الجهاز وذلك عند مستوى الأهمية $\alpha = 0.20$.

الحل: لنشكل الفرضية الابتدائية H_0 بالشكل : $H_o : \sigma^2 = 4$

ثم نشكل الفرضية البديلة H_1 بالشكل : $H_1 : \sigma^2 \neq 4$

لنوجد الآن إحصاء الاختبار Z بالشكل: $S^2 = \frac{1}{n-1} \sum \left(X_i - \overline{X} \right)^2 = 10.57$

ثم نوجد إحصاء الاختبار: $\chi^2 = \frac{(n-1) S^2}{\sigma_0^2} = \frac{2(10.57)}{4} = 5.29$

لدينا $\alpha = 0.20 \Leftarrow \chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0.10}(2) = 0.211$ ، $\chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0.90}(2) = 4.61$

نلاحظ أنّ: $\chi^2_{0.90}(2) = 4.61 < 5.29 = \chi^2$ أي أنّنا نرفض الفرضية H_0 عند مستوى الأهمية $\alpha = 0.20$.

مثال: استخدم بيانات المثال السابق في اختبار نفس الفرضية H_o عند مستوى الأهمية $\alpha = 0.10$.

الحل : لدينا من المثال السابق أنّ إحصاء الاختبار $\chi^2 = 5.29$.

لنوجد الآن قيمة χ^2 عند مستوى الأهمية $\alpha = 0.10$:

$$\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0.05}(2) = 0.103 \quad , \quad \chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0.95}(2) = 5.99$$

بالمقارنة نجد أنّ : $\chi^2 = 5.29 > 5.99$ وأنّ $\chi^2 = 5.29 \not< 0.103$ وبالتالي نستنتج أنّ المعلومات الإحصائية المرجوة في هذا النص لا تقدم دلالة كافية لرفض الفرضية H_0 .

ثامناً – اختبار نسبة تبايني مجتمعين طبيعيين :

تعود أهمية المقارنة بين تبايني مجتمعين طبيعيين إلى كثرة استخدام التباين في معرفة خصائص المجتمعات الإحصائية . وغالباً ما نرغب في مقارنة دقة جهاز للقياس بدقة جهاز آخر ، واستقرار كل خط من خطين للإنتاج في استخلاص مادة اليورانيوم .

مبرهنة (8): لتكن $X_1, X_2, \ldots, X_m$ عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu_1, \sigma_1^2)$ ، ولتكن $Y_1, Y_2, \ldots, Y_n$ عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N(\mu_2, \sigma_2^2)$.

آ – لاختبار الفرضية: $H_o : \sigma_1^2 = \sigma_2^2$ مقابل الفرضية: $H_1 : \sigma_1^2 \neq \sigma_2^2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$\frac{S_2^2}{S_1^2} \begin{cases} \leq F_{\frac{\alpha}{2}}(n-1\,,\,m-1) \\ \text{أو} \\ \geq F_{1-\frac{\alpha}{2}}(n-1\,,\,m-1) \end{cases}$$

ب – لاختبار الفرضية: $H_o : \sigma_1^2 = \sigma_2^2$ مقابل الفرضية: $H_1 : \sigma_1^2 > \sigma_2^2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$\frac{S_2^2}{S_1^2} \geq F_{1-\alpha}(n-1,\, m-1)$$

ج – لاختبار الفرضية: $H_o : \sigma_1^2 = \sigma_2^2$ مقابل الفرضية $H_1 : \sigma_1^2 < \sigma_2^2$ فإنّنا نرفض الفرضية H_o إذا كان:

$$\frac{S_2^2}{S_1^2} \le F_{\alpha}(n-1,\ m-1)$$

مثال: اعتمدت دراسة عن معالجات معينة ، وكان هدف هذه الدراسة هو التحليل المخبري لتعيين دواء الدايوكسين وهو دواء مهم وشائع الاستخدام وهو سام أيضاً. ومن المعروف أيضاً أنّ مستوى الجرعات للمرضى الذين يتناولونه فوق سن الـ 64 يجب أنّ تكون أقل منها لدى المرضى الأكثر شباباً . وقد أجريت الدراسة لتبيأنّ النماذج المستقلة المأخوذة من كل مجموعة ، وقد تم تحديد مستوى جرعات الدايوكسين لدى كل مريض تم اختياره ، ونتيجة الدراسة حصلنا على المعلومات الآتية :

المرضى الذين أعمارهم 64 وما دون	المرضى الذين أعمارهم فوق 64
$n = 41$ ملغ / يوم	$m = 29$ ملغ / يوم
$\overline{Y} = 0.265$ ملغ / يوم	$\overline{X} = 0.268$ ملغ / يوم
$S_2 = 0.102$ ملغ / يوم	$S_1 = 0.068$ ملغ / يوم

والمطلوب :

اختبر الفرضية: $H_o : \sigma_1^2 = \sigma_2^2$ ضد الفرضية: $H_1 : \sigma_1^2 \neq \sigma_2^2$ وذلك عند مستوى المعنوية $\alpha = 0.05$.

الحل: أنّ هذا الاختبار هو اختبار باتجاهين ، لذا فإنّ الفرضية H_0 يجب أنّ نرفضها لصالح الفرضية H_1 البديلة إذا كانت قيمة إحصاء الاختبار كبيرة جداً أو صغيرة جداً . ولندرس الحالة التي يكون فيها المجتمعان طبيعيين . أنّ منطقة الرفض هي :

$$\frac{S_2^2}{S_1^2} \begin{cases} \le F_{\frac{\alpha}{2}}(n-1\ ,\ m-1) \\ \text{أو} \\ \ge F_{1-\frac{\alpha}{2}}(n-1\ ,\ m-1) \end{cases}$$

لنوجد إحصاء الاختبار والقيمتين المعياريتين ثم نقارن بينهما .

$$\frac{S_2^2}{S_1^2} = \left(\frac{0.102}{0.068}\right)^2 = 2.26$$

ومن جدول توزيع فيشر نجد : $F_{1-\frac{\alpha}{2}}(n-1, m-1) = F_{0.975}(40, 28) = 2.05$

بالمقارنة نجد أنّ : $\frac{S_2^2}{S_1^2} = 2.26 > 2.05 = F_{0.975}(40, 28)$

وبالتالي نرفض الفرضية H_0 عند مستوى الأهمية $\alpha = 0.05$.

سادساً – بعض الاختبارات غير الوسيطية:

1 – اختبار الفرق بين متوسطي عينتين غير مستقلتين:

استخدمنا في بعض الفقرات السابقة اختبار الفرق بين عينتين مستقلتين ولكن قد يحدث في الحياة العملية أنّ المشاهدات تكون عبارة عن أزواج مرتبة من الأعداد ، كان يكون هناك علاقة بين طول الشخص ووزنه فإذا أخذنا عينة من هذه الأزواج المرتبة بالشكل : $(x_1, y_1)\ ,\ (x_2, y_2)\ , \ldots,\ (x_n, y_n)$. ونرغب في إجراء اختبار للمقارنة بين مجموعتي القراءتين لنفس المفردات في العينة في المرة الأولى والثانية وذلك من أجل استنتاج تأثير المؤثر على مفردات العينة وفي هذه الحالة ندرس الفرق بين مقدارين غير مستقلين عن بعضهما البعض وهما متوسط القراءة الأولى ومتوسط القراءة الثانية. وفي هذه الحالة سوف نستخدم توزيع ستيودنت وذلك كما يلي:

آ – نوجد الفرق $d_i = x_i - y_i$, ثم نوجد متوسط هذه الفروق $\bar{d}$.

ب – نوجد الانحراف المعياري لـ $\bar{d}$, ثم نوجد الإحصاء $T = \dfrac{\bar{d}}{\sigma_{\bar{d}} / \sqrt{n}}$.

ج – نقارن قيمة T مع قيمة t المعيارية ، فإذا كان : $T > t_{1-\frac{\alpha}{2}}(n-1)$.

نرفض الفرضية H_o ونقبل الفرضية البديلة H_1 .

مثال: يمثل الجدول التالي درجات سبعة من الطلاب في مادتي الجبر والتحليل:

رقم الطالب	1	2	3	4	5	6	7
X درجة الجبر	62	84	77	57	62	90	82
y درجة التحليل	53	75	65	55	67	85	79

والمطلوب اختبار فيما إذا كانت هناك فروق معنوية بين متوسطي درجات مادتي الجبر والتحليل أي لنختبر الفرضية $H_o : \bar{d} = 0$ ضد الفرضية البديلة $H_1 : \bar{d} \neq 0$ عند مستوى المعنوية $\alpha = 0.05$ لسهولة الحل نحاول حساب الإحصاء T من الجدول التالي :

رقم الطالب	X درجة الجبر	Y درجة التحليل	$d = x - y$	$d - \bar{d}$	$(d - \bar{d})^2$
1	62	53	9	4	16
2	84	75	9	4	16
3	77	65	12	7	49
4	57	55	2	− 3	9

5	62	67	− 5	− 10	100
6	90	85	5	0	0
7	82	79	3	− 2	4
المجموع	−	−	**35**	**0**	**194**

حيث $\bar{d} = \dfrac{\sum d}{n} = \dfrac{35}{7} = 5$.

لنوجد الآن الانحراف المعياري لـ $\bar{d}$ وذلك بالشكل :
$$\sigma_{\bar{d}}^2 = \frac{\sum\left(d_i - \bar{d}\right)^2}{n-1} = \frac{194}{6} = 32.3$$

وبالتالي فإنّ : $\sigma_{\bar{d}} = 5.686$

لنوجد الآن إحصاء الاختبار T بالشكل :
$$T = \frac{\bar{d}}{\sigma_{\bar{d}} / \sqrt{n}} = \frac{5}{\frac{5.686}{\sqrt{7}}} = 2.327$$

لدينا عدد البيانات 7 وبالتالي درجة حرية هذا التوزيع هي 6 أي : $t_{1-\frac{\alpha}{2}}(n-1) = t_{0.975}(6) = 2.447$

نلاحظ أنّ : $t_{0.975}(6) = 2.447 > |T| = 2.327$. وبالتالي لا نستطيع رفض الفرضية H_o القائلة أنّه لا توجد فروق ذات دلالة بين درجات الطلاب في مادتي الجبر والتحليل .

2 – اختبار بارتليت:

إذا أخذنا مجموعة **r** من العينات من مجتمعات طبيعية مستقلة عن بعضها البعض ، وكانت تباينات هذه العينات هي $S_1^2, S_2^2, \dots, S_r^2$ وأحجامها $n_1, n_2, \dots, n_r$ على الترتيب . وإذا أردنا اختبار الفرضية $H_o : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2$ حيث σ_i^2 هو تباين المجتمع **i** ضد الفرضية H_1 القائلة بأنّ تباينين على الأقل غير متساويين، فإنّنا نرفض الفرضية H_0 عند مستوى الأهمية α إذا كان :
$$B > \chi^2_{1-\alpha}(r-1)$$

حيث :
$$B = \frac{1}{C}\left\{(n-r)\,Ln\left[\frac{1}{n-r}\sum_{i=1}^{r}(n_i - 1)S_i^2\right] - \sum_{i=1}^{r}(n_i - 1)\,Ln\,S_i^2\right\}$$

$$C = 1 + \frac{1}{3(r+1)}\left\{\sum_{i=1}^{r}\left(\frac{1}{n_i - 1}\right) - \frac{1}{n-r}\right\} \;;\; n = \sum_{i=1}^{r} n_i$$

مثـال: أخذت أربـع عينات من أربعة مجتمعات طبيعية مستقلة عن بعضـها البعض ورتبت أحجام هذه العينات وتبايناتها في الجدول التالي :

تباين المجتمع	تباين العينة	حجم العينة
σ_1^2	0.47	7
σ_2^2	0.4614	7
σ_3^2	0.5714	7
σ_4^2	0.1995	7

استخدم اختبار بارتليت في اختبار الفرضية الابتدائية: تساوي تباينات المجتمعات الأربعة مقابل الفرضية البديلة: اثنان على الأقل مختلفان عند مستوى الأهمية $\alpha = 0.01$.

الحل :

أنّ منطقة الرفض هي $B > \chi^2_{1-\alpha}(r-1)$ (*)

نقوم الآن بترتيب الحسابات في الجدول التالي وذلك من أجل سهولة إيجاد B ومقارنتها مع القيمة المعيارية.

$(n_i-1)\,Ln\,S_i^2$	$Ln\left(S_i^2\right)$	$(n_i-1)\,S_i^2$	S_i^2	n_i-1	المجتمع
− 4.5301	− 0.755	2.82	0.47	6	1
− 4.6409	− 0.7734	2.7684	0.4614	6	2
− 3.3579	− 0.5596	3.4282	0.5714	6	3
− 9.6716	− 1.6119	1.197	0.1995	6	4
− 22.2007		10.2138		24	المجموع

$$C = 1+\frac{1}{9}\left(\frac{2}{3}-\frac{1}{24}\right) = 1.0694$$

$$B = \frac{24\,Ln\left(\frac{10.2138}{24}\right) + 22.2007}{1.0694} = \frac{24\,Ln\,(0.4256) + 22.2007}{1.0694}$$

$$= \frac{24\,(-0.8543) + 22.2007}{1.0694} = \frac{-20.5021 + 22.2007}{1.0694}$$

$$= \frac{1.6986}{1.0694} = 1.588$$

ومن جدول توزيع كاي مربع نجد: $\chi^2_{1-\alpha}(r-1) = \chi^2_{0.99}(3) = 11.34$

بالمقارنة نجد أنّ : $B = 1.588 < 11.34 = \chi^2_{0.99}(3)$

أي أنّنا نقبل الفرضية H_9 عند مستوى الأهمية $\alpha = 0.01$ والتي تعني أنّ تباينات المجتمعات الأربعة متساوية .

3 – اختبار الاستقلال:

في كثير من الأحيان ، نقوم بتصنيف مجموعة من المشاهدات وفق أسلوبين (تصنيفين) أو أكثر ، فمثلاً قد نصنف طلاب جامعة البعث وفق الجنس من جهة ووفق الجنسية من جهة أخرى ووفق المستوى التعليمي في المرحلة الثانوية من جهة ثالثة... ففي مثل هذه الحالات يكون مفيداً جداً معرفة العلاقة بين الجنس والمستوى التعليمي في المرحلة الثانوية أو معرفة العلاقة بين جنسية الطالب والمستوى التعليمي . يسمى الاختبار الذي يدرس مثل هذه الحالات باختبار الاستقلال أو باختبار كاي مربع .

مبرهنة: إذا صنفت مجموعة من المشاهدات وفق أسلوبين أحدهما X والآخر Y في جدول توافق مدخلاته O_{ij} وأردنا اختبار الفرضية H_0 القائلة بأنّ X , Y مستقلان مقابل الفرضية H_1 القائلة بأنّهما غير مستقلين ، فإنّنا نرفض الفرضية H_0 إذا كان : $U^2 = \sum_{i,j} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} > \chi_{1-\alpha}(r-1)(c-1)$ عند مستوى الأهمية α .

حيث r هو عدد الأسطر ، c هو عدد الأعمدة في جدول التوافق .

مثال: ألقينا حجر نرد تسعين مرة وكان التوزيع التكراري لظهور الأوجه الستة هي كما في الجدول التالي :

الوجه الظاهر	1	2	3	4	5	6
عدد مرات ظهوره	15	7	15	8	30	15

والمطلوب اختبار فيما إذا كان حجر النرد متوازن أم لا ؟ .

الحل : نحن نعلم أنّه عندما نرمي حجر النرد تسعين مرة فإنّ كل وجه سوف يظهر بنفس الاحتمال وبالتالي يكون عدد مرات ظهور كل وجه هو 15 وهذا هو عدد المرات المتوقعة أو النظرية لظهور أي وجه (e_{ij}) .

ولإجراء اختبار كاي مربع نتبع الخطوات التالية :

1– نحدد عدد المرات المشاهدة لظهور أي رقم (O_{ij}), ثم عدد المرات المتوقعة لظهور أي رقم (e_{ij}) .

2– نحسب الفرق بين القراءة المشاهدة والقراءة المتوقعة أي نحسب الفرق $(O_{ij} - e_{ij})$ ، ثم نربع هذا الفرق .

3– نقسم $(O_{ij} - e_{ij})^2$ على القراءة المتوقعة في كل قراءة , ثم نحسب الإحصاء $U^2 = \sum \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$.

4- نرفض الفرضية H_o إذا كان $U^2 > \chi^2_{1-\alpha}(r-1)(c-1)$.

يمكننا إيجاد قيمة الإحصاء U^2 مباشرة من العلاقة أو من الجدول التالي :

رقم الوجه	1	2	3	4	5	6	المجموع
O_{ij}	15	7	15	8	30	15	90
e_{ij}	15	15	15	15	15	15	90
$O_{ij} - e_{ij}$	0	− 8	0	− 7	15	0	0
$\left(O_{ij}-e_{ij}\right)^2$	0	64	0	49	225	0	338
$\frac{\left(O_{ij}-e_{ij}\right)^2}{e_{ij}}$	0	4.3	0	3.3	15	9	22.6

أي أنّ $U^2 = 22.6$. ولدينا من جدول كاي مربع : $\chi^2_{1-\alpha}(r-1)(c-1) = \chi^2_{0.95}(5) = 11.07$

لأنه لدينا ست خلايا تحت شرط واحد وهو مجموع قراءاتها يساوي تسعين ، نلاحظ أنّ :

$$U^2 = 22.6 > 11.07 = \chi^2_{0.95}(5)$$

وبالتالي فإنّنا نرفض الفرضية H_o أي أنّ حجر النرد غير متزن .

مثال: أراد فريق طبي أنّ يتعرف فيما إذا كانت هناك علاقة بين الدم وشدة الإصابة بمرض معين ، لأجل هذا قام الفريق الطبي بدراسة 1500 حالة وصنفها في الجدول التالي ، والمسمى جدول التوافق :

نوع الدم / شدة المرض	A	B	AB	O	**المجموع**
بسيط	543	211	90	476	1320
متوسط	44	22	8	31	105
شديد	28	9	7	31	75
المجموع	**615**	**242**	**105**	**538**	**1500**

والمطلوب :

آ – أوجد احتمال أنّ يكون دمه من النوع **A** .

ب – أوجد احتمال أنّ تكون شدة المرض بسيطة .

ج – أوجد احتمال أنّ يكون دمه من النوع **A** وشدة مرضه بسيطة .

الحل :

آ - P (أنّ يكون دمه من النوع A) $= \frac{615}{1500} = 0.41$

ب - P (أنّ تكون شدة المرض بسيطة) $= \frac{1320}{1500} = 0.88$

ج - P (أنّ يكون دمه من النوع A وشدة مرضه بسيطة) $= \frac{543}{1500} = 0.362$

لنفرض الآن أنّ شدة المرض ونوع الدم مستقلين عن بعضهما البعض ، فعندئذٍ يكون احتمال أنّ تكون الإصابة بسيطة والدم من النوع A يساوي حاصل ضرب احتمالهما أي :

$$(0.88)\ (0.41) = 0.3608$$

نلاحظ أنّ هذا الاحتمال لا يساوي تماماً قيمة الاحتمال المحسوب في (ج)، ولكنه قريب منه . ويعني هذا الاحتمال أنّنا نتوقع أنّ يكون عدد الأشخاص الذين تكون إصاباتهم بسيطة ودمهم من النوع A يساوي :

$$(0.3608)\ (1500) = 541.2$$

في حين أنّ عدد الأشخاص الذين شاهدهم الفريق الطبي في هذه الفئة يساوي 543 . نلاحظ أيضاً أنّ هذين الاحتمالين متقاربان .

تقوم فلسفة اختبار كاي مربع للاستقلال على إيجاد مقياس لهذه الأخطاء الناجمة عن تقريب القيم المشاهدة بالقيم المتوقعة بغرض الاستقلال وهذا المقياس هو الإحصاء التالي :

$$U^2 = \sum_{i,j} \left(O_{ij} - e_{ij}\right)^2 / e_{ij}$$

حيث : O_{ij} يمثل عدد المشاهدات في السطر i و العمود j .

e_{ij} تمثل عدد الأفراد المتوقع تواجدهم في السطر i والعمود j وبفرض الاستقلال .

نلاحظ من المثال السابق أنّ : $e_{ij} = \frac{(n_{i\bullet})(n_{\bullet j})}{n}$

حيث : $n_{i\bullet}$ تمثل مجموع المشاهدات في السطر i .

$n_{\bullet j}$ تمثل مجموع المشاهدات في العمود j .

n يمثل عدد المشاهدات الكلي في جدول التوافق .

ولمعرفة فيما إذا كانت فرضية الاستقلال مقبولة أم لا فإنّنا نقدر فيما إذا كانت U^2 صغيرة أم كبيرة . فإذا كانت U^2 كبيرة فإنّ هذا يعني أنّ الأخطاء كبيرة وبالتالي فالاستقلال مرفوض ، وأما إذا كانت U^2 صغيرة فإنّ هذا

يعني أنّ الأخطاء صغيرة وبالتالي فالاستقلال مقبول ، ولمعرفة صغر أو كبر U^2 نقارنها مع القيمة المعيارية لكاي مربع أي أنّنا نرفض الفرضية H_0 إذا كان : $U^2 > \chi^2_{1-\alpha}\,(r-1)\,(c-1)$

حيث r هو عدد الأسطر ، c هو عدد الأعمدة في جدول التوافق .

وبالعودة إلى المثال السابق واستخدام بياناته فإنّنا نستطيع أنّ نبين أنّ قيم e_{ij} يمكن ترتيبها في الجدول التالي :

نوع الدم / شدة المرض	A	B	AB	O
بسيط	541.2	212.96	92.40	473.44
متوسط	43.05	16.94	7.35	37.66
شديد	30.75	12.10	5.25	26.9

لنوجد الآن الإحصاء U^2 بالشكل :

$$U^2 = \frac{(543-541.2)^2}{541.2} + \frac{(211-212.96)^2}{212.96} + \dots + \frac{(31-26.90)^2}{26.90} = 5.12$$

ومن جدول توزيع كاي مربع نجد : $\chi^2_{1-\alpha}(r-1)(c-1) = \chi^2_{0.95}(6) = 12.592$

بالمقارنة نجد أنّ : $U^2 = 5.12 > 12.592 = \chi^2_{0.95}(6)$

أي أنّنا لا نستطيع أنّ نرفض الفرضية H_0 وبالتالي نجد أنّ لا علاقة بين نوع الدم وشدة المرض .

4 – اختبار حسن المطابقة:

تعتمد الدراسات الإحصائية على فرض نموذج إحصائي معين تخضع له المشاهدات المتوفرة ، فمثلاً نقول المتغير العشوائي الذي يحكم إحدى التجارب التي أعيدت مئة مرة هو متغير حداني أو طبيعي ، وقد نقول أنّ متوسط الذكاء لمجموعة من الأطفال في مدينة حمص يخضع للتوزيع الطبيعي أو قد نقول أنّ عدد حوادث السير التي تحدث عند تقاطع طريق معين تخضع لتوزيع بواسون ، فكيف يمكننا التأكد من أنّ المشاهدات المتوفرة تخضع لنموذج إحصائي معين . ولمعرفة مثل هذه النماذج يمكننا استخدام اختبار كاي مربع لحسن المطابقة والذي يعتبر أحد الاختبارات الهامة المعدة لهذا الغرض .

1– اختبار حسن المطابقة للتوزيع الحداني :

مبرهنة: لنفترض أنّه لدينا جدول توزيع تكراري لتجربة ما تكراراته O_i وفئاته E_i حيث $i = 1 , 2 , \dots , r$ وأردنا اختبار الفرضية القائلة بأنّ المتغير X الممثل بهذا الجدول يخضع لنموذج إحصائي معين تماماً فإنّنا نرفض هذه الفرضية H_0 إذا كانت :

$$U^2 = \sum_{i=1}^{r} \frac{(O_i - e_i)^2}{e_i} > \chi^2_{1-\alpha}(r-1)$$

حيث : $e_i = Np(X \in E_i)$, $O_i \geq 5$ لكل i .

ملاحظات هامة :

آ – عند حل أي تمرين باستخدام اختبار حسن المطابقة يجب التأكد أولاً من أنّ $O_i \geq 5$ لكل i ، فإذا كانت $O_i < 5$ لبعض قيم i فإنّنا نقوم بدمج الفئة التي تكرارها أقل من 5 مع الفئة السابقة لها أو الفئة اللاحقة لها ثم نقوم بتطبيق المبرهنة أو القانون المتعلق بالاختبار المفروض ونلاحظ أنّ r هو عدد الأسطر في الجدول الناتج بعد عملية الدمج .

ب – إذا كان النموذج الإحصائي غير معين تعييناً تاماً مثل X يخضع لتوزيع بواسون الذي وسيطه λ فهذا غير محدد تحديداً تاماً وذلك لأنّ التوزيع يتغير بتغير الوسيط λ . لهذا نقوم أولاً بتقدير قيمة الوسطاء المجهولة باستخدام طريقة العزوم مثلاً وبعدها تطبق المبرهنة المتعلقة بالاختبار المفروض مع تغير درجات الحرية من $(r-1)$ إلى $(r-k-1)$ حيث k هو عدد الوسطاء التي تم تقديرها .

مثال: إذا فرضنا أنّ الجدول التالي يمثل عدد العائلات التي لديها خمسة أطفال وبفرض أنّ X يدل على عدد الذكور حيث عدد العائلات كان 320 عائلة :

عدد الأطفال الذكور	0	1	2	3	4	5
عدد العائلات	8	40	88	110	56	18

والمطلوب اختبار فيما إذا كانت هذه البيانات تخضع لتوزيع حداني .

الحل :

آ – إذا كان النموذج الإحصائي الذي يمثل هذه البيانات هو التوزيع الحداني $b\left(5, \frac{1}{2}\right)$ صحيحاً فإنّه بالإمكان حساب احتمالات جميع الأحداث المتعلقة بتلك التجربة باستخدام قانون التوزيع الحداني مباشرة حيث $n=5$, $P=\frac{1}{2}$.

$$P(X=0)=\frac{1}{32} \quad ; \quad P(X=1)=\frac{5}{32} \quad ; \quad P(X=2)=\frac{10}{32}$$

$$P(X=3)=\frac{10}{32} \quad ; \quad P(X=4)=\frac{5}{32} \quad ; \quad P(X=5)=\frac{1}{32}$$

ب – نوجد القيم المتوقعة لكل من الأحداث E_i من العلاقة التالية: $e_i = Np(X \in E_i)$ وعدد العائلات

المتوقعة والتي لديها ذكر واحد يساوي: $50 = \frac{5}{32} \times 320$ وهكذا ، ويمكننا ترتيب النتائج في الجدول التالي:

E_i	O_i	e_i
0	8	10
1	40	50
2	88	100
3	110	100
4	56	50
5	18	10

ج – نوجد من هذا الجدول إحصاء الاختبار U^2 من العلاقة :

$$U^2 = \sum \frac{\left(O_{ij} - e_{ij}\right)^2}{e_i} = \frac{(8-10)^2}{10} + \frac{(40-50)^2}{50} + \dots + \frac{(18-10)^2}{10} = 11.96$$

د – نوجد القيمة المعيارية $\chi^2_{1-\alpha}(r-1)$ من جدول كاي مربع ، فإذا فرضنا أنّ $\alpha = 0.05$ ولدينا عدد الصفوف r يساوي 6 فنجد: $\chi^2_{1-\alpha}(r-1) = \chi^2_{0.95}(5) = 11.07$

هـ – نقارن إحصاء الاختبار U^2 مع القيمة المعيارية $\chi^2_{1-\alpha}(r-1)$ فنجد أنّ:

$$U^2 = 11.96 > 11.07 = \chi^2_{0.95}(5)$$

وبالتالي فإنّنا نرفض الفرضية H_0 القائلة بأنّ هذه البيانات تخضع لتوزيع حداني $b\left(5, \frac{1}{2}\right)$.

2 – اختبار حسن المطابقة للتوزيع الطبيعي :

بنفس الطريقة التي اتبعناها في حساب حسن المطابقة في التوزيع الحداني يمكن اختبار حسن مطابقة التوزيع لبعض المشاهدات التي تواجهنا ويختلف حساب درجات الحرية عن التوزيع السابق ، لأنّه لابد من تقدير متوسط وتباين العينة في كل مرة وكذلك تحديد مجموع التكرارات ، أي أنّه توجد ثلاثة شروط في حالة التوزيع الطبيعي .

مثال: لتكن لدينا مجموعة المشاهدات ممثلة بالجدول التكراري التالي :

الفئات	أقل من 15	15–20	20–25	25–30	30–35	أكبر من 35
التكرارات	3	7	15	20	9	4

والمطلوب اختبار فيما إذا كانت المشاهدات تخضع للتوزيع الطبيعي .

الحل :

آ – لحساب ذلك نوجد أولاً المتوسط الحسابي $\overline{X}$ والانحراف المعياري S لهذه المشاهدات ، ونحن نعلم أنّ التوزيع الطبيعي يتحدد بمعرفة متوسطه μ وتباينه σ^2 ، ولما كانت μ و σ^2 مجهولتين فإنّنا نقدرهما من الجدول كما يأتي:

$$\hat{\mu} = \overline{X} = \frac{\sum_i O_i \ X_i}{\sum_i O_i} = 25.7$$ حيث X_i هي مراكز الفئات ، O_i هي التكرارات .

وكذلك : $$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_i \left(X_i - \overline{X}\right)^2 = 37.6996 \Rightarrow \hat{\sigma} = S = 6.14$$

ب – نعين الحدود العليا للفئات ومن ثم نوجد القيم المعيارية لها ولتكن Z ومن ثم نوجد الاحتمالات المقابلة لها من العلاقة $P(Z \le z)$ ونحسب من ذلك الاحتمال والتكرار المتوقع لتلك الفئة كما في الجدول التالي :

الفئات	أقل من 15	15–20	20–25	25–30	30–35	أكبر من 35
الحد الأعلى للفئة	15	20	25	30	35	–
القيمة المعيارية للحد الأعلى للفئة	− 1.74	− 0.93	− 0.11	0.70	1.51	–
$P(Z \le z)$	0.0409	0.1762	0.4562	0.758	0.9345	1
الاحتمال P	0.0409	0.1353	0.28	0.3018	0.1765	0.0655
التكرار المتوقع $P.\sum O_1$	2.4	7.9	16.2	17.5	10.2	3.8

ولتوضيح طريقة الحساب فإنّ القيم المعيارية نحصل عليها من العلاقة: $Z = \dfrac{X - \overline{X}}{\sigma}$

فمثلاً القيمة المعيارية للفئة الأولى نحصل عليها كما يلي: $Z = \dfrac{15 - 25.7}{6.14} = -1.74$

وكذلك الأمر فإنّ القيمة المعيارية للفئة الثانية نحصل عليها من العلاقة: $Z = \dfrac{20 - 25.7}{6.14} = -0.93$

وهكذا ... أما السطر الرابع وهو الاحتمالات فنجدها مباشرة من جدول التوزيع الطبيعي المعياري وأما السطر الخامس فنجد قيمه كما يلي :

الاحتمال الأول $= 0.0409$

الاحتمال الثاني $= 0.1762 - 0.0409 = 0.1353$

الاحتمال الثالث $= 0.4562 - 0.1762 = 0.28$

وهكذا ....

والسطر الأخير نحصل عليه بضرب الاحتمال في مجموع التكرارات 58 .

نلاحظ أنّه لابد من دمج الفئتين الأولى والثانية وكذلك الفئتين الخامسة والسادسة لأنّ القيمة المتوقعة في الفئة الأولى والأخيرة أقل من 5 . لذا نقوم بدمج هذه الفئات فنحصل على الجدول التالي :

O_i	10	15	20	13
e_i	10.3	16.2	17.5	14

وبالتالي نجد :

$$U^2 = \sum \frac{(O_i - e_i)^2}{e_i} = \frac{(10-10.3)^2}{10.3} + \frac{(15-16.2)^2}{16.2} + \frac{(20-17.5)^2}{17.5} + \frac{(13-14)^2}{14} = 0.5$$

لنوجد الآن درجة حرية توزيع كاي مربع بالشكل :

بما أنّه توجد لدينا أربع خلايا وكذلك تقدير قيمة كل من المتوسط والانحراف المعياري والالتزام بشرط مجموع التكرارات فإنّ عدد درجات الحرية هي : $r - k - 1 = 4 - 2 - 1 = 1$ وبالتالي فإنّ :

$$\chi^2_{1-\alpha}(r-k-1) = \chi^2_{0.95}(1) = 3.84$$

بالمقارنة نجد أنّ : $U^2 = 0.53 > 3.84| = \chi^2_{0.95}(1)$

أي أنّنا نقبل الفرضية H_0 بأنّ المشاهدات المعطاة لها التوزيع الطبيعي.

3 – اختبار حسن المطابقة لتوزيع بواسون :

يعتبر توزيع بواسون من التوزيعات الهامة في دراسة العديد من الظواهر العشوائية ، حيث يلعب دوراً أساسياً في العديد من المسائل الفيزيائية ومسائل التحليل المخبري وله تطبيقات واسعة في كثير من الظواهر العشوائية مثل انبعاث الغازات الإشعاعية من طاقة كوكب نووي وانبعاث الإلكترونات من مادة حساسة تحت تأثير الضوء وعدد الذرات التي تفنى في الثانية من مادة مشعة . وأيضاً له تطبيقات في مجالات بحوث العمليات والعلوم الإدارية . وبصورة خاصة فإنّ توزيع بواسون كثير الفائدة في تفتيش المنتجات المصنعة ومراقبتها وتصنيفها .

تكون المشاهدات أو التوزيعات الفعلية المشاهدة في توزيع بواسون عادة معطاة سواء من التجارب أو من أي ظاهرة طبيعية . وكل ما نفعله هو إيجاد توزيع بواسون المناظر ومن ثم إيجاد القيم المتوقعة نظرياً ، ومن ثم نستخدم الإحصاء U^2 المعرف سابقاً . كما ونجد أنّ لتوزيع بواسون حالة واحدة حيث يمكن توليد توزيع بواسون النظري إذا علم متوسطه ومجموع التكرارات الكلي ، وبالتالي يوجد شرطان في كل استخدامات توزيع بواسون لحسن المطابقة .

مثال: لتكن لدينا مجموعة من المشاهدات ممثلة بالجدول التكراري التالي :

القيمة	0	1	2	3	4	5	6 أو أكثر
التكرار	19	26	27	13	11	2	0

والمطلوب اختبار فيما إذا كانت هذه البيانات تخضع لتوزيع بواسوني .

الحل :

آ – نلاحظ أنّ : $\overline{X} = \frac{\sum f \cdot x}{\sum f} = \frac{173}{98} = 1.765$

ب – $P(X=x) = \frac{(1.765)^x}{x!} e^{-1.765} \quad , \quad x = 0, 1, 2, \dots, 6$

ج – لنوجد القيم المتوقعة من العلاقة : $e_i = N P(X=x)$

فنجد : $e_0 = 16.78 \quad ; \quad e_1 = 29.61 \quad ; \quad e_2 = 26.13$

$e_3 = 15.37 \quad ; \quad e_4 = 6.78 \quad ; \quad e_{(\text{5أو أكبر})} = 3.33$

ويمكن إيجاد القيمة الأخيرة (5 أو أكبر) من العلاقة: $e_{(\text{5أو أكبر})} = 98 - e_0 - e_1 - e_2 - e_3 - e_4 = 3.33$ وبما أنّ القيمة الأخيرة أصغر من 5 وبالتالي لابد من دمجها مع القيمة السابقة لها وبالتالي نحصل على الجدول التالي:

القيمة	0	1	2	3	4 أو أكبر
القيمة المشاهدة	19	26	27	13	13
القيمة المتوقعة	16.8	29.61	26.1	15.37	10.1

د – لنوجد الإحصاء U^2 بالشكل :

$$U^2 = \sum_{i=1}^{r} \frac{(O_i - e_i)^2}{e_i}$$

$$= \frac{(19-16.8)^2}{16.8} + \frac{(26-29.61)^2}{29.61} + \frac{(27-26.1)^2}{26.1} + \frac{(13-15.37)^2}{15.37} + \frac{(13-10.1)^2}{10.1} = 1.964$$

هـ – وبما أنّ عدد الخلايا هو 5 وعدد الشروط 2 فإنّ درجة الحرية هي 3 وبالتالي فإنّ القيمة المعيارية لتوزيع كاي مربع تصبح بالشكل: $\chi^2_{0.95}(3) = 7.81$

و – نلاحظ أنّ: $U^2 = 1.964 \triangleright 7.81 = \chi^2_{0.95}(3)$.

وبالتالي نقبل الفرضية H_0 والتي تقول بأنّ المشاهدات تتبع توزيع بواسون .

سابعاً – توطئة نيمان بيرسون:

تتعلق الفرضيات الإحصائية بسلوك متغيرات عشوائيّة قابلة للملاحظة فلو فرضنا أنّ $X_1, X_2, \dots, X_n$ متغيّرات عشوائية فيمكن التعبير عنها كنقطة $X = (X_1, X_2, \dots, X_n)$ في فضاء العينة E وهو فضاء اقليدي ذو n بعداً. لتكن C منطقة من فضاء العيّنة وبما أنّ لكل متغير عشوائي توزيع احتمالي فيمكننا من الناحية النظرية على الأقل أنّ نحسب احتمال أنّ تنتمي النقطةX إلى المنطقة C أي : $P(X \in C)$ ونقول أنّ أية فرضية تتعلق بسلوك متغيرات عشوائية هي فرضية إحصائية.

لتكن $f(x_1, x_2, \dots, x_n, \theta)$ الكثافة المشتركة للمتغيرات العشوائية $X_1, X_2, \dots, X_n$ حيث θ وسيط ذو n بعداً يأخذ قيمه في فضاء الوسيط Ω والحالة التي نواجهها غالباً في حركة كون المتغيرات العشوائية $X_1, X_2, \dots, X_n$ عيّنة عشوائيّة وتكون الكثافة المشتركة عندئذ عبارة عن جداءات الكثافات $f(x_i, \theta)$ أي أنّ:

$$f(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = L(\underline{x}, \theta) = L(\theta)$$

ويتحدد الموقف من الفرضية المطروحة ونرمز لها بالفرضّية الابتدائية H_0 وفقا لقيم الوسيط θ لذلك نقسّم فضاء الوسيط Ω إلى مجموعتين جزئيتين منفصلتين $\Omega_{H_0}, \Omega_{H_1}$ حيث أنّ : $\Omega = \Omega_{H_0} \cup \Omega_{H_1}$ كما نقول أنّ الفرضيّة صحيحة إذا كانت $\theta \in \Omega_{H_0}$ وغير صحيحة إذا كانت $\theta \in \Omega_{H_1}$ وإذ ا تضمنت المجموعة الجزئيّة Ω_{H_0} نقطة واحدة من الفضاء Ω نقول أنّ الفرضية H_0 هي فرضية بسيطة وإلا فإنّنا نقول أنّ الفرضية الإحصائية هي فرضيّة مركبة , ثم نتّخذ القرار فيما إذا كانت الفرضية H_0 مقبولة أو مرفوضة اعتماداً على النتائج التي نستخلصها من العيّنة العشوائية المأخوذة من المجتمع الإحصائي $f(x, \theta)$ وبعبارة أدق: نقسم فضاء العيّنة E إلى منطقتين $C, \grave{C}$ بحيث أنّ : $E = C \cup \grave{C}$ أو $\grave{C} = E \backslash C$ وهاتان المنطقتان مستقلتان عن الوسيط θ , فنرفض الفرضية H_0 إذا كان $X = (X_1, X_2, \dots, X_n) \in C$ ونقبلها إذا كان $X = (X_1, X_2, \dots, X_n) \in \grave{C}$ وقبول الفرضية يعني أنّنا لم نجد وفقا للبيان الإحصائي الذي توفره لنا العينة دلالات كافية لرفض العيّنة حيث تسمى القاعدة أو الطريقة التي تحدد لنا منطقة الرفض بالاختبار الإحصائي أي أنّ عملية اختبار منطقة الرّفض هي الاختبار الإحصائي . وعند اتخاذ قرار الرفض أو القبول يمكن أنّ نرتكب أحد الخطأين , والاختبار الأمثل هو الاختبار الذي يجعل كلاًّ من الخطأين أصغر ما يمكن ,إلا أنّه من أجل حجم معّين ومحدد تؤدي كل حيطة نتخذها ضد أحد الخطأين إلى ثغرة في حيطتنا ضد الخطأ الأخر لأنّه لا يمكننا دفع مقداري الخطأين في اتجاه واحد بحيث يزدادان أو ينقصان معاّ أي لأنّه يمكننا التحكّم بخطأ واحد فقط لذلك نعرّف طريقة الاختبار C بحيث يكون: $P[(X_1, X_2, \dots, X_n) \in C \backslash H_0]$ أي أنّ الحل الأمثل هو أنّ نختار C بحيث يكون الخطأ الأول محدود والخطأ الثاني اصغر ما يمكن .

نص توطئة نيمان بيرسون:

لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي قانونه $f(x,\theta)$ ولتكن θ', θ'' قيمتين ثابتتين ومحدودتين لـ θ حيث أنّ: $\Omega = \{\theta; \theta = \theta', \theta = \theta''\}$ وليكن k عددا موجباً ولتكن C مجموعة جزئية من فضاء العينة بحيث يكون :

$a)\ \frac{L(\theta')}{L(\theta'')} \le k \qquad for\ \ \underline{x} = (x_1, x_2, \dots, x_n) \in C$

$b)\ \frac{L(\theta')}{L(\theta'')} \ge k \qquad for\ \ \underline{x} = (x_1, x_2, \dots, x_n) \in C'$

$c)\ \alpha = p[(X_1, X_2, \dots, X_n) \in C;\ H_0]$ حيث أنّ C' هي متممة C

عندئذ تكون C منطقة رفض مثلى حجمها α وذلك من أجل اختبار الفرضية الابتدائية: $H_0: \theta = \theta'$ مقابل الفرضية البديلة: $H_1: \theta = \theta''$.

وسوف ندرس ثلاثة نماذج من هذه التوطئة:

النموذج الأول: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه μ مجهول و σ^2 معلوم. وبفرض أنّ μ_0 عدداً ثابتاً فأوجد منطقة الرفض المثلى لاختبار الفرضية الابتدائية: $H_0: \mu = \mu_0$ مقابل الفرضية البديلة $H_1: \mu = \mu_1 > \mu_0$

الحل: لتحديد منطقة الرفض المثلى نستخدم توطئة نيمان بيرسون وتكون منطقة الرفض المثلى هي المنطقة C المحددة بالعلاقة: $C = \left\{(x_1, x_2, \dots, x_n); \frac{L(\theta')}{L(\theta'')} \le k\right\}$ حيث يتم تعيين k من العلاقة التالية:

$\alpha = p[(X_1, X_2, \dots, X_n) \in C;\ H_0]$ علماً أنّ H_0 صحيحة. وبالتالي فإنّ منطقة الرفض المثلى تصبح بالشكل : $C = \{(x_1, x_2, \dots, x_n); \sum x_i \ge a\}$

أما إذا كانت الفرضية البديلة من الشكل : $H_1: \mu = \mu_1 < \mu_0$ فإنّنا نحصل بطريقة مشابهة على منطقة الرفض المثلى من الشكل : $C = \{(x_1, x_2, \dots, x_n); \sum x_i \le a\}$.

النموذج الثاني: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه $\mu = 0$ و σ^2 مجهول. وبفرض أنّ σ_0^2 عدداً ثابتاً فأوجد منطقة الرفض المثلى لاختبار الفرضية الابتدائية: $H_0: \sigma^2 = \sigma_0^2 = 1$ مقابل الفرضية البديلة $H_1: \sigma^2 = \sigma_0^2 = 2$ موافق لـ $\alpha = 0.05$.

الحل: أنّ منطقة الرفض المثلى هي المجموعة : C= $\left\{(x_1, x_2, \dots, x_n); \frac{L(1)}{L(2)} \le k\right\}$

حيث تتعين k من العلاقة: $p\left[(X_1, X_2, \dots, X_n) \in \text{C};\ \text{صحيحة}\ H_0\right] = \alpha$

وبحساب النسبة نجد أنّ منطقة الرفض المثلى هي: C= $\{(x_1, x_2, \dots, x_n): \sum_{i=1}^{n} x_i^2 \ge h\}$

النموذج الثالث: إذا لم تكن المتغيرات العشوائية $X_1, X_2, \dots, X_n$ مستقلة وعلينا إيجاد منطقة الرفض المثلى

لاختبار الفرضية الابتدائية : $H_0: f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)$

ضد الفرضية البديلة : $H_1: f(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n)$

وباستخدام توطئة نيمان بيرسون نجد أنّ منطقة الرفض المثلى هي المجموعة:

$$C = \left\{(x_1, x_2, \dots, x_n); \frac{g(x_1,x_2,\dots,x_n)}{h(x_1,x_2,\dots,x_n)} \le k\right\}$$

حيث نعين k من العلاقة : $p\left[(X_1, X_2, \dots, X_n) \in C; \text{صحيحة } H_0\right] = \alpha$

مثال: لتكن X_1, X_2 عيّنة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي قانونه:

$$f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \ , \ x > 0 \ , \ \theta > 0$$

والمطلوب أوجد منطقة الرفض المثلى لاختبار الفرضية: $H_0: \theta = 2$ ضد الفرضية البديلة: $H_1: \theta = 4$ وذلك عند مستوى الأهمية $\alpha = 0.05$.

الحل: أنّ منطقة الرفض المثلى هي المجموعة: $C = \left\{(x_1, x_2); \frac{L(\theta')}{L(\theta'')} \le k\right\}$ حيث أنّ k ثابت موجب نحدده من العلاقة: $p\left[(X_1, X_2) \in C\ ;\ \text{صحيحة } H_0\right] = 0.05$.

$$C = \left\{(x_1, x_2); 4e^{-\left(\frac{x_1+x_2}{4}\right)} \le k\right\} = \left\{(x_1, x_2); -\left(\frac{x_1+x_2}{4}\right) \le \mathrm{Ln}\left(\frac{k}{4}\right)\right\}$$

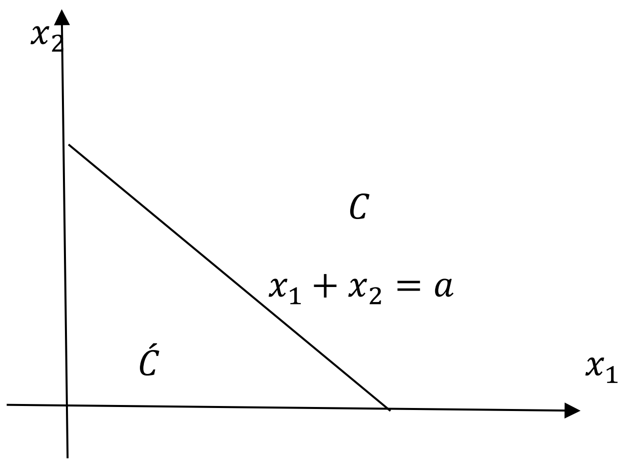
أي أنّ:

$$C = \left\{(x_1, x_2); x_1 + x_2 \ge -4\mathrm{Ln}\left(\frac{k}{4}\right)\right\}$$

لنحدد الآن الثابت k بحيث يكون الخطأ الأول محدود وهو $\alpha = 0.05$:

$$P[(x_1, x_2) \in C; \theta = 2\,] = P[x_1 + x_2 \ge a\,; \theta = 2]$$

نلاحظ أنّ الكثافة من أجل x_2 و x_1 هي موجبة دوما مما يعني أنّ المنطقة C موجودة في الربع الأول لذا نرسم المستقيم $x_1 + x_2 = a$ فتكون المنطقة C هي المنطقة المحددة بالشكل التالي :



لنحسب الاحتمال : $P[x_1+x_2 \geq a \backslash \theta=2] = P[(x_1,x_2) \in C \backslash \theta=2] =$

$$\int_0^{\infty}\int_0^{\infty}\frac{1}{4}e^{-\left(\frac{x_1+x_2}{2}\right)}dx_1dx_2 = 1-\int_0^{\infty}\int_0^{\infty}\frac{1}{4}e^{-\left(\frac{x_1+x_2}{2}\right)}dx_1dx_2$$

$$= 1-\frac{1}{4}\int_0^{k}e^{-\frac{x_1}{2}}\left(\int_0^{k-x_1}e^{-\frac{x_2}{2}}dx_2\right)dx_1 = const$$

نضع هذا الثابت الأخير مساوياً لـ $\alpha = 0.05$ وبحل هذه المعادلة الآسية نحصل على المطلوب.

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي دالة كثافته الاحتمالية $f(x)$

ولاختبار الفرضية الابتدائية : $H_0: f(x) = \frac{e^{-1}}{x!}$ علماً أنّ $x = 0,1,2,\dots$.

مقابل الفرضية البديلة : $H_1: f(x) = \left(\frac{1}{2}\right)^{x+1}$ علماً أنّ $x = 0,1,2,\dots$.

الحل:

أنّ منطقة الرفض المثلى هي المجموعة: $C=\left\{(x_1,x_2,\dots,x_n); \frac{g(x_1,x_2,\dots,x_n)}{h(x_1,x_2,\dots,x_n)} \leq k\right\}$ بحيث يتم تعيين k

من العلاقة : $p\left[(X_1,X_2,\dots,X_n) \in C; \text{صحيحة } H_0\right] = \alpha$.

لدينا:

$$\frac{g(x_1,x_2,\dots,x_n)}{h(x_1,x_2,\dots,x_n)} = \frac{\prod_{i=1}^{n}\left[\frac{e^{-1}}{(x_i)!}\right]}{\prod_{i=1}^{n}\left(\frac{1}{2}\right)^{x_i+1}} = \frac{\frac{e^{-n}}{(x_1)!(x_2)!\dots(x_n)!}}{\left(\frac{1}{2}\right)^{n}\left(\frac{1}{2}\right)^{x_1+x_2+\dots+x_n}} = \frac{(2e^{-1})^{n}2^{\sum_{x=1}^{n}x_i}}{\prod_{x=1}^{n}(x_i)!} \leq k$$

إذا كان $k > 0$ فإنّ منطقة الرفض المثلى التي تعينها مجموعة النقاط $(x_1,x_2,\dots,x_n)$ تصبح بالشكل :

$$C=\{(x_1,x_2,\dots,x_n); (\textstyle\sum_{x=1}^{n}x_i)Ln2 - Ln[\prod_{x=1}^{n}(x_i)!] \leq Lnk - n\,Ln(2e^{-1}) = c\}$$

ومن أجل $n=1$, $k=1$ فإنّ المتباينة السابقة تصبح بالشكل:

$$\frac{g(x_1,x_2,\dots,x_n)}{h(x_1,x_2,\dots,x_n)} = \frac{2^{x_1}}{(x_1)!} \leq \frac{e}{2}$$

وبالتالي فإنّ منطقة الرفض المثلى تكون محققة من أجل مجموعة نقاط المجموعة:

$C=\{x_1; x_1 = 0,3,4,5,\dots\}$.

وتكون قوة الاختبار عندما تكون الفرضية الابتدائية H_0 صحيحة بالشكل:

$$p(X_1 \in C;\ H_0) = 1-p(X_1=1,2;H_0) = 1-\left(\frac{1}{e}+\frac{1}{2e}\right) = 1-0.552 = 0.448$$

وتكون قوة الاختبار عندما تكون الفرضية الابتدائية H_1 صحيحة بالشكل:

$$p(X_1 \in C;\ H_1) = 1-p(X_1=1,2;H_1) = 1-\left(\frac{1}{4}+\frac{1}{8}\right) = 1-0.375 = 0.625$$

جامعة دمشق	امتحان مقرر الإحصاء الرياضي	الاســـم : ....
كلية العلوم	لطلاب السنة الثالثة قسم الرياضيات	الدرجة : مائـــــة
قسم الرياضيات	الفصل الثاني للعام الدراسي 2012 – 2013	المـدة : ساعتان

السؤال الأول (20 درجة) :

لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي دالة كثافته $f(x, \theta)$. أثبت أنّ الشـرط اللازم والكـافي لكي يكون الإحصاء $Y_1 = U_1(X_1, X_2, \ldots, X_n)$ إحصاءً كافياً بالنسبة لـ θ هو أنّ يتحقق الشرط التالي:

$$f(x_1, x_2, \ldots, x_n, \theta) = \underbrace{g_1(y_1, \theta)}_{Y_1 \text{ دالة الكثافة الهامشية لـ}} \underbrace{H(x_1, x_2, \ldots, x_n)}_{\theta \text{ لا يتعلق بـ}}$$

السؤال الثاني (25 درجة) :

ليكن X, Y متغيرين عشوائيين بواسونين مستقلين وسيطاهما على الترتيب λ_1, λ_2 , ولنفرض أنّ:

$U = X + Y \quad , \quad V = Y$ متغيرين عشوائيين والمطلوب:

أوجد دالة الكثافة المشتركة $g(u, v)$ للمتغيرين العشوائيين U, V ثم أوجد دالة الكثافة $g(u)$ للمتغير U واستنتج نوعه.

السؤال الثالث (20 درجة) :

لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي له التوزيع الطبيعي وسيطاه $\mu = 0 , \sigma^2 = \theta$.

استخدم متباينة كرامير – راو في إثبات أنّ التقدير $Y = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ يبلغ الحد الأدنى للتباين .

السؤال الرابع (17 درجة) :

أخذت عينة عشوائية حجمها $n = 100$ ، من مجتمع طبيعي ، تباينه $\sigma^2 = 25$. فإذا كان متوسط هذه العينة هو 56.18 . أوجد 95 % فترة ثقة لمتوسط هذا المجتمع الطبيعي μ , علماً أنّ القيمة المعيارية المناسبة هي 1.96 .

السؤال الخامس (18 درجة) :

لتكن X_1 , X_2 عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي دالة كثافته معرفة بالعلاقة التالية :

$$f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad , x > 0 \; , \; \theta > 0$$

والمطلوب:

استخدم توطئة نيمان - بيرسون في إيجاد منطقة الرفض المثلى في اختبار الفرضية الابتدائية: $H_0 : \theta = 3$ ضد الفرضية البديلة: $H_1 : \theta = 5$.

انتهت الأسئلة

ملاحظة : يسمح باستخدام الآلات الحاسبة

دمشق في 18 / 6 / 2013 — مدرس المقرر: د. سلطان محمد الصلخدي

سلم تصحيح مقرر الإحصاء الرياضي

لطلاب السنة الثالثة قسم الرياضيات

الفصل الثاني للعام الدراسي 2012 – 2013

الجواب الأول (20 درجة) :

1 – لزوم الشرط : لنفرض أنّ الشرط (1) محقق ولنثبت أنّ Y_1 هو إحصاء كافي بالنسبة لـ θ .

لتكن $i = 2,3,\dots,n$, $Y_i = U_i(X_1, X_2, \dots, X_n)$, إحصاءات اختيارية وبالتالي فإنّ الكثافة المشتركة لهذه المتغيرات العشوائية تعطى بالشكل: $g(y_1, y_2, \dots, y_n, \theta) = f(x_1, x_2, \dots, x_n, \theta)|J|$ وذلك بتبديل $x_1, x_2, \dots, x_n$ بقيمها من التحويلات المماثلة.

$$g(y_1, y_2, \dots, y_n, \theta) = f\left[u_1^{-1}\left(\underline{y}, \theta\right), u_2^{-1}\left(\underline{y}, \theta\right), \dots, u_n^{-1}\left(\underline{y}, \theta\right), \theta\right]|J|$$

وبما أنّ الشرط (1) محقق فإنّ:

$$g(y_1, y_2, \dots, y_n, \theta) = g_1(y_1, \theta)H(x_1, x_2, \dots, x_n)|J|$$

ولكن الكثافة الشرطية h تعطى بالعلاقة:

$$h(y_2, y_3, \dots, y_n \backslash y_1, \theta) = \frac{g(y_1, y_2, \dots, y_n, \theta)}{g_1(y_1, \theta)}$$
$$= \frac{g_1(y_1, \theta)H(x_1, x_2, \dots, x_n)|J|}{g_1(y_1, \theta)} = H(x_1, x_2, \dots, x_n)|J|$$

وبما أنّ H مستقل عن θ وكذلك معين اليعقوبي, فإنّ الكثافة الشرطية لا تتعلق بـ θ وبالتالي يكون Y_1 إحصاءً كافياً لـ θ.

2 – كفاية الشرط: نفرض أنّ Y_1 إحصاء كافي بالنسبة لـ θ ولنثبت أنّ الشرط (1) محقق:

بما أنّ Y_1 إحصاء كافي بالنسبة لـ θ فإنّ:

$$g(y_1, y_2, \dots, y_n, \theta) = g_1(y_1, \theta)h(y_2, y_3, \dots, y_n \backslash y_1)$$

لكن:

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta).f(x_2, \theta)\dots f(x_n, \theta) = g(y_1, y_2, \dots, y_n, \theta)|J^*|$$

حيث: J^* هو التحويل المعاكس من X إلى Y وبالتالي:

$$f(x_1, x_2, \dots, x_n, \theta) = g_1(y_1, \theta)h(y_2, y_3, \dots, y_n \backslash y_1)|J^*|$$

حيث: J^*, h غير تابع لـ θ, أي أنّ:

$$f(x_1, x_2, \dots, x_n, \theta) = g_1(y_1, \theta)H(x_1, x_2, \dots, x_n)$$

أي أنّ الشرط (1) محقق.

الجواب الثاني (25 درجة) :

بما أنّ X,Y متغيران عشوائيان مستقلان فإنّ دالة الكثافة المشتركة تعطى بالعلاقة:

$$f(x,y)=f(x).f(y)=\frac{\lambda_1^x}{x!}e^{-\lambda_1}.\frac{\lambda_2^y}{y!}e^{-\lambda_2}$$

أي أنّ :
$$f(x,y)=\frac{\lambda_1^x.\lambda_2^y}{x!y!}e^{-(\lambda_1+\lambda_2)}$$

ومنه نستطيع أنّ نجري التحويل التالي: $v=y \quad ; \quad u=x+y\} \Rightarrow x=u-v \quad ,y=v$

وبالتالي فإنّ دالة الكثافة المشتركة لـ U,V تعطى بالشكل التالي:

$$g(u,v)=\frac{\lambda_1^{u-v}\lambda_2^v}{(u-v)!.v!}e^{-(\lambda_1+\lambda_2)} \quad ,(u,v)\in B$$

علماً بأنّ:
$$B=\{(u,v):v=0,1,2,\dots,u \quad ; \quad u=0,1,2,\dots\}$$

لنوجد دالة الكثافة للمتغير U كما يلي:

$$g(u)=\sum_{v=0}^{u}g(u,v)=\frac{e^{-(\lambda_1+\lambda_2)}}{u!}\sum_{v=0}^{u}\frac{u!}{(u-v)!.v!}\lambda_1^{u-v}\lambda_2^v$$

وبالتالي:
$$g(u)=\frac{(\lambda_1+\lambda_2)^u}{u!}e^{-(\lambda_1+\lambda_2)} \quad : \quad u=0,1,\dots.$$

وهو توزيع بواسوني وسيطه$(\lambda_1+\lambda_2)$ وبذلك يتم المطلوب.

الجواب الثالث (20 درجة):

لدينا دالة الكثافة للتوزيع الطبيعي معرفة بالشكل:

$$f(x,\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{1}{2}\frac{x^2}{\theta}}$$

وبأخذ اللوغاريتم الطبيعي للطرفين نجد:

$$Lnf(x,\theta)=-\frac{1}{2}Ln(2\pi)-\frac{1}{2}Ln(\theta)-\frac{1}{2}\frac{x^2}{\theta}$$

وباشتقاق الطرفين جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial Lnf(x,\theta)}{\partial\theta}=-\frac{1}{2\theta}+\frac{x^2}{2\theta^2}$$

وبالاشتقاق مرة أخرى جزئياً بالنسبة لـ θ نجد:

$$\frac{\partial^2 Lnf(x,\theta)}{\partial\theta^2}=\frac{1}{2\theta^2}-\frac{x^2}{\theta^3}$$

لنوجد الآن معلومات فيشر من العلاقة:

$$I=E\left(-\frac{\partial^2 Lnf(x,\theta)}{\partial\theta^2}\right)=E\left(\frac{X^2}{\theta^3}-\frac{1}{2\theta^2}\right)=\frac{1}{\theta^3}E(X^2)-E\left(\frac{1}{2\theta^2}\right)$$
$$=\frac{1}{\theta^3}Var(X)-\frac{1}{2\theta^2}=\frac{1}{\theta^3}(\theta)-\frac{1}{2\theta^2}=\frac{1}{\theta^2}-\frac{1}{2\theta^2}=\frac{1}{2\theta^2}$$

وبالتالي فإنّ الحد الأدنى للتباين كما تحدده متباينة كرامير – راو يساوي $\frac{2\theta^2}{n}$.

لنوجد الآن تباين المقدِّر $T=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ بالشكل:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i^2\right)$$
$$=\frac{1}{n^2}\sum_{i=1}^{n}Var\left(X_i^2\right)=\frac{1}{n^2}\left[n\left(Var\left(X^2\right)\right)\right]=\frac{1}{n}Var\left(X^2\right)$$

وذلك لأنّ المتغيرات العشوائية $X_1,X_2,\dots,X_n$ مستقلة , ولكن تباين المتغير X^2 يحسب من العلاقة التالية:

$$Var(X^2)=E(X^4)-[E(X^2)]^2=(\mu_4-\mu_2^2)$$
$$=\left[\left(\frac{4!}{(2^2)(2!)}\right)\theta^2-\theta^2\right]=3\theta^2-\theta^2=2\theta^2$$

علماً أنّ عزم المجتمع من المرتبة الرابعة حول الصفر للمتغير الطبيعي يعطى من العلاقة التالية:

$\mu_{2r}=\frac{(2r)!}{2^r(r!)}\theta^2$, حيث لدينا في هذا المثال أنّ : $\sigma^2=\theta$.

وبالتالي فإنّ:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{2\theta^2}{n}$$

بالموازنة بين تباين المقدِّر والحد الأدنى للتباين كما تحدده متباينة كرامير – راو نجد أنّهما متساويان وبالتالي فإنّ هذا المقدِّر هو مقدِّر غير منحاز وذو تباين أصغري.

الجواب الرابع (17 درجة) :

لدينا : $\frac{\alpha}{2}=0.025 \Rightarrow \quad \alpha=0.05 \Rightarrow \quad 1-\alpha=0.05 \Rightarrow$

$$1-\frac{\alpha}{2}=0.975 \Rightarrow z_{1-\frac{\alpha}{2}}=z_{0.975}=1.96$$

أنّ فترة الثقة تعطى بالشكل :

$$C.I=\left[\overline{X}-\frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}}\quad;\quad\overline{X}+\frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}}\right]$$
$$=\left[56.18-\left(\frac{5}{10}\right)(1.96)\quad;\quad 56.18+\left(\frac{5}{10}\right)(1.96)\right]=[55.20\quad;\quad 57.16]$$

الجواب الخامس (18 درجة) :

أنّ منطقة الرفض المثلى هي المجموعة: $C=\left\{(x_1,x_2);\frac{L(\theta')}{L(\theta'')}\le k\right\}$ حيث أنّ k ثابت موجب نحدده من العلاقة: $p\left[(X_1,X_2)\in C;\text{ صحيحة } H_0\right]=0.05$.

لنوجد الآن $L(\theta')=L(3)$ ثم $L(\theta'')=L(5)$ ثم نأخذ النسبة كما يلي:

$$L(3)=\frac{1}{9}e^{-\left(\frac{x_1+x_2}{3}\right)}$$

$$C=\left\{(x_1,x_2);4e^{-\left(\frac{x_1+x_2}{4}\right)}\leq k\right\}=\left\{(x_1,x_2);-\left(\frac{x_1+x_2}{4}\right)\leq \mathrm{Ln}(\frac{k}{4})\right\}$$

أي أنّ:

$$C=\left\{(x_1,x_2);x_1+x_2\geq -4\mathrm{Ln}(\frac{k}{4})\right\}$$

$$L(2)=2{x_1}^{2-1}.2{x_2}^{2-1}=4x_1x_2$$

لنأخذ النسبة :

$$\frac{L(\theta')}{L(\theta'')}=\frac{L(3)}{L(5)}=\frac{1}{4x_1x_2}$$

وبالتالي فإنّ منطقة الرفض المثلى ما هي إلا مجموعة النقاط :

$$C=\left\{(x_1,x_2):\frac{1}{4x_1x_2}\leq k\right\}=\left\{(x_1,x_2):x_1x_2\geq \frac{1}{4k}\right\}$$

انتهت الأجوبة

دمشق في 18 / 6 / 2013

مدرس المقرر: د. سلطان محمد الصلخدي

جامعة دمشق	**امتحان مقرر الإحصاء الرياضي**	**الاســم : ....**
كلية العلوم	**لطلاب السنة الثالثة قسم الرياضيات**	**الدرجة : مائـــة**
قسم الرياضيات	**الدورة الاستثنائية للعام الدراسي 2012 – 2013**	**المـدة : ساعتان**

السؤال الأول (25 درجة) :

ليكن X متغيراً عشوائياً له التوزيع الطبيعي المعياري وليكن Y متغيراً عشوائيّاً أخراً له توزيع كاي مربّع درجة حريته n ولنفرض أنّ X, Y مستقلان والمطلوب ايجاد دالة الكثافة الاحتمالية للمتغير $T = \frac{X}{\sqrt{\frac{y}{n}}}$.

السؤال الثاني (30 درجة) :

ليكن X, Y متغيرين عشوائيين دالة كثافتهما المشتركة معرفة بالعلاقة :

$$f_{X,Y}(x,y) = \frac{2}{\theta^2} e^{-\frac{x+y}{\theta}} \qquad ; 0 < x < y < \infty \quad , \theta > 0$$

والمطلوب:

1 – أوجد دالة الكثافة الهامشية لكل من Y و X .

2 – أوجد $E(Y)$ و $Var(Y)$.

3 – بفرض أنّ : $E(Y/x) = \varphi(x)$ أثبت أنّ $Var(\varphi(x)) \le Var(Y)$.

السؤال الثالث (25 درجة) :

إذا كانت 41 , 39 , 31 , 29 , 35 , 23 , 20 عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu_1, \sigma^2)$ وكانت 34 , 25 , 45 , 37 43 عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N(\mu_2, \sigma^2)$. أوجد 95 % فترة ثقة للفرق $(\mu_1 - \mu_2)$, وذلك بفرض أنّ القيمة المعيارية هي (2.228) .

السؤال الرابع (20 درجة) :

لتكن $X_1, X_2, \dots, X_{16}$ عينة عشوائية مأخوذة من مجتمع طبيعي بحيث أنّ $\bar{X} = 80$, $S^2 = 25$.

والمطلوب: اختبر الفرضية الابتدائية: $H_0 : \mu = 75$ ضد الفرضية البديلة : $H_1 : \mu < 75$.

علماً أنّ القيمة المعيارية هي (2) .

انتهت الأسئلة

ملاحظة : يسمح باستخدام الآلات الحاسبة

دمشق في 18 / 8 / 2013 — **مدرس المقرر: د. سلطان محمد الصلخدي**

سلم تصحيح مقرر الإحصاء الرياضي

لطلاب السنة الثالثة قسم الرياضيات

الدورة الاستثنائية للعام الدراسي 2012 – 2013

الجواب الأول (25 درجة) :

نلاحظ أنّ: $Y\sim\chi^2(n)$; $X\sim N(0,1)$ و أنّ X,Y مستقلان .

نفرض وجود متغير مساعد $Z=Y$ فنحسب دالة الكثافة المشتركة لـ T,Z أولا ومن ثم دالة الكثافة لـ T :

إذاً خطوات الحل هي :

1)– إيجاد الكثافة الاحتمالية المشتركة لـ T,Z .

2)– إيجاد الكثافة الهامشية لـ T .

$X=T\sqrt{\frac{y}{n}}$, $Y=Z$ ومنه يصبح لدينا:

$$f(t,y)=f[u_1(t,y),u_2(t,y)].|J|=f\left(x=t\sqrt{\frac{y}{n}},y=y\right).|J|$$

$$=\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]\left[\frac{1}{(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(y)^{\frac{n}{2}-1}\mathrm{e}^{-\frac{y}{2}}\right].|\mathrm{J}|$$

$$=\frac{1}{\sqrt{2\pi}(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(y)^{\frac{n}{2}-1}e^{-\frac{t^2y}{2n}}e^{-\frac{y}{2}}|\mathrm{J}|$$

نحسب معين اليعقوبي بالشكل الآتي:

$$\mathrm{J}=\begin{vmatrix}\frac{\partial x}{\partial t} & \frac{\partial x}{\partial y}\\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial y}\end{vmatrix}=\begin{vmatrix}\sqrt{\frac{y}{n}} & \text{مقدارما}\\ 0 & 1\end{vmatrix}=\sqrt{\frac{y}{n}}$$

إذا أصبح لدينا:

$$f(t,y)=\frac{1}{\sqrt{2\pi}(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{y}{2}\left(1+\frac{t^2}{n}\right)}(y)^{\frac{n}{2}-1}\left|\sqrt{\frac{y}{n}}\right|$$

$$=\frac{1}{\sqrt{n}\sqrt{2\pi}(2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{y}{2}(1+\frac{t^2}{n})}y^{\frac{n-1}{2}}\quad ;\quad y>0\,;t\in R$$

وبالتالي فإنّ:

$$f(t)=\int_0^\infty f(t,y)dy=\int_0^\infty \frac{1}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{y}{2}(1+\frac{t^2}{n})}y^{\frac{n-1}{2}}dy$$

نجري تغيير في المتحول بالشكل:

$$z=\frac{y}{2}\left(1+\frac{t^2}{n}\right)\Longrightarrow y=\frac{2z}{(1+\frac{t^2}{n})}\Longrightarrow dy=\frac{2dz}{(1+\frac{t^2}{n})}$$

ومنه نجد :

$$f(t)=a\int_0^\infty e^{-z}\left(\frac{2z}{1+\frac{t^2}{n}}\right)\frac{2dz}{(1+\frac{t^2}{n})}=\frac{a2^{\frac{n+1}{2}}}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}}\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz$$

علما أنّ: $. a=\frac{1}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)}$

وبالتالي يمكن أنّ نكتب:

$$f(t)=\frac{2^{\frac{n+1}{2}}}{\sqrt{n\pi}(2)^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)(1+\frac{t^2}{n})^{\frac{n+1}{2}}}\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz$$

ومن دالة غاما نجد: $\Gamma(\vartheta)=\int_0^\infty x^{\vartheta-1}e^{-x}dx\Longrightarrow\frac{n-1}{2}=\vartheta-1\Longrightarrow\vartheta=\frac{n+1}{2}$

أي أنّ: $\int_0^\infty z^{\frac{n-1}{2}}e^{-z}dz=\Gamma(\frac{n+1}{2})$

نعوض مباشرةً فنجد أنّ :

$$f(t)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)(1+\frac{t^2}{n})^{\frac{n+1}{2}}}=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}$$

وهي دالة الكثافة الاحتمالية لتوزيع ستيودنت درجة حريته n .

الجواب الثاني (30 درجة) :

$$f_X(x)=\int_{y=x}^{y=\infty}f_{X,Y}(x,y)dy=\int_{y=x}^{y=\infty}\frac{2}{\theta^2}e^{-\frac{x+y}{\theta}}dy$$
$$=\frac{2}{\theta}e^{-\frac{x}{\theta}}\int_{y=x}^{y=\infty}\frac{1}{\theta}e^{-\frac{y}{\theta}}dy=\frac{2}{\theta}e^{-\frac{x}{\theta}}\left[-e^{-\frac{y}{\theta}}\right]_{y=x}^{y=\infty}$$
$$=\frac{2}{\theta}e^{-\frac{x}{\theta}}\left[e^{-\frac{x}{\theta}}\right]=\frac{2}{\theta}e^{-\frac{2x}{\theta}}\qquad;\theta>0,x>0$$

أي أنّ X أسي وسيطه $\lambda=\frac{2}{\theta}$.

$$f_y(y)=\int_{x=0}^{x=y}f_{X,Y}(x,y)dx=\int_{x=0}^{x=y}\frac{2}{\theta^2}e^{-\frac{x+y}{\theta}}dx$$
$$=\frac{2}{\theta}e^{-\frac{y}{\theta}}\int_{x=0}^{x=y}\frac{1}{\theta}e^{-\frac{x}{\theta}}dy=\frac{2}{\theta}e^{-\frac{y}{\theta}}\left[-e^{-\frac{x}{\theta}}\right]_{x=0}^{x=y}$$
$$=\frac{2}{\theta}e^{-\frac{y}{\theta}}\left[1-e^{-\frac{y}{\theta}}\right]=\frac{2}{\theta}e^{-\frac{y}{\theta}}-\frac{2}{\theta}e^{-\frac{2y}{\theta}}\qquad;\theta>0,y>0$$

لنوجد التوقع الرياضي:

$$E(Y)=\int_0^\infty y f_y(y)dy=\int_0^\infty y\left[\frac{2}{\theta}e^{-\frac{y}{\theta}}-\frac{2}{\theta}e^{-\frac{2y}{\theta}}\right]dy=\frac{2}{\theta}\underbrace{\int_0^\infty y\,e^{-\frac{y}{\theta}}dy}_{I_1}-\frac{2}{\theta}\underbrace{\int_0^\infty y\,e^{-\frac{2y}{\theta}}dy}_{I_2}$$

$$I_1=\int_0^\infty y\,e^{-\frac{y}{\theta}}dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{1}{\theta}y=\mathfrak{z}\Rightarrow dy=\theta d\mathfrak{z}$$

وبالتالي فإنّ :

$$I_1=\theta^2\int_0^\infty \mathfrak{z}\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^2\int_0^\infty \mathfrak{z}^{2-1}\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^2\Gamma(2)=\theta^2$$

حساب التكامل الثاني:

$$I_2=\int_0^\infty y\,e^{-\frac{2y}{\theta}}dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{2}{\theta}y=t\Rightarrow dy=\frac{\theta}{2}dt$$

وبالتالي فإنّ :

$$I_2=\frac{\theta^2}{4}\int_0^\infty t\,e^{-t}dt=\frac{\theta^2}{4}\int_0^\infty t^{2-1}\,e^{-t}dt=\frac{\theta^2}{4}\Gamma(2)=\frac{\theta^2}{4}$$

$$E(y)=2.\frac{1}{\theta}.\theta^2-2.\frac{1}{\theta}.\theta^4=2\theta-\frac{1}{2}\theta=\frac{3}{2}\theta\ Var(y)=\underbrace{Ey^2}_{?}-[E(y)]^2$$

$$E(y^2)=\int_0^\infty y^2 f_y(y)dy=\int_0^\infty y^2\left[\frac{2}{\theta}e^{-\frac{y}{\theta}}-\frac{2}{\theta}e^{-\frac{2y}{\theta}}\right]dy$$

$$=\frac{2}{\theta}\underbrace{\int_0^\infty y^2\,e^{-\frac{y}{\theta}}dy}_{I_1}-\frac{2}{\theta}\underbrace{\int_0^\infty y^2\,e^{-\frac{2y}{\theta}}dy}_{I_2}$$

$$I_1=\int_0^\infty y^2\,e^{-\frac{y}{\theta}}dy$$

$$\frac{1}{\theta}y=\mathfrak{z}\Rightarrow dy=\theta d\mathfrak{z}\Rightarrow I_1=\theta^3\int_0^\infty \mathfrak{z}^2\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^3\int_0^\infty \mathfrak{z}^{3-1}\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^3\Gamma(3)=2\theta^3$$

$$I_2=\int_0^\infty y^2\,e^{-\frac{2y}{\theta}}dy$$

$$\frac{2}{\theta}y=t\Rightarrow dy=\frac{\theta}{2}dt\Rightarrow I_2=\frac{\theta^3}{8}\int_0^\infty t^2\,e^{-t}dt=\frac{\theta^3}{8}\int_0^\infty t^{3-1}\,e^{-t}dt=\frac{\theta^3}{8}\Gamma(3)=\frac{\theta^3}{4}$$

$$E(y^2)=\frac{2}{\theta}\,2\theta^3-\frac{2}{\theta}\,\frac{\theta^3}{4}=4\theta^2-\frac{1}{2}\theta^2=\frac{7}{2}\theta^2$$

$$Var(y)=Ey^2-[E(y)]^2=\frac{7}{2}\theta^2-\left(\frac{3}{2}\theta^2\right)^2=\frac{5}{4}\theta^2$$

$$f_{Y/X}(y/x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{\frac{2}{\theta^2}e^{-\frac{x+y}{\theta}}}{\frac{2}{\theta}e^{-\frac{2x}{\theta}}}=\frac{1}{\theta}\,e^{-\frac{1}{\theta}(y-x)}\,;\begin{cases}0<x<y<\infty\\ \quad 0<x\end{cases}$$

$$E(y/x) = \int_{y=x}^{y=\infty} y\, f_{Y/X}(y/x)dy = \int_{y=x}^{y=\infty} y \left[\frac{1}{\theta}\, e^{-\frac{1}{\theta}(y-x)}\right] dy$$
$$= \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{y=x}^{y=\infty} y\, e^{-\frac{1}{\theta}y} dy = \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} \theta^2\, \mathfrak{z}\, e^{-\mathfrak{z}} d\mathfrak{z}$$
$$= \theta\, e^{\frac{x}{\theta}} \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} \mathfrak{z}\, e^{-\mathfrak{z}} d\mathfrak{z}$$

نكامل بالتجزئة لذلك نفرض :

$$\left.\begin{matrix} u = \mathfrak{z} \Rightarrow du = d\mathfrak{z} \\ dv = e^{-\mathfrak{z}} d\mathfrak{z} \Rightarrow v = -\, e^{-\mathfrak{z}} \end{matrix}\right\} \Rightarrow E(y/x) = \theta\, e^{\frac{x}{\theta}} \left[-\mathfrak{z}\, e^{-\mathfrak{z}} \Big|_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} + \int_{\mathfrak{z}=\frac{x}{\theta}}^{\mathfrak{z}=\infty} e^{-\mathfrak{z}} d\mathfrak{z} \right]$$
$$= \theta\, e^{\frac{x}{\theta}} \left[\frac{x}{\theta}\, e^{-\frac{x}{\theta}} + e^{-\frac{x}{\theta}}\right] = x + \theta = \varphi(x)$$

$$Var(\varphi(x)) = \underbrace{E[\varphi(x)]^2}_{?} - \underbrace{(E[\varphi(x)])^2}_{?}$$

$$E[\varphi(x)] = E[x+\theta] = E[x] + \theta = \frac{\theta}{2} + \theta = \frac{3}{2}\theta$$
$$E[\varphi(x)]^2 = E[x+\theta]^2 = E[x^2 + 2\theta x + \theta^2]$$
$$= E[x^2] + 2\theta E[x] + \theta^2$$
$$= E[x^2] + 2\theta\left[\frac{\theta}{2}\right] + \theta^2$$
$$= \frac{\theta^2}{2} + \theta^2 + \theta^2 = \frac{5}{2}\theta^2$$

حيث لدينا X أسي وسيطه $\lambda = \frac{2}{\theta}$.

$$\Rightarrow Var(X) = E[X^2] - [E(X)]^2 \Rightarrow$$
$$E[X^2] = Var(X) + [E(X)]^2$$
$$= \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2 = \frac{\theta^2}{4} + \frac{\theta^2}{4} = \frac{\theta^2}{2} \Rightarrow$$
$$Var(\varphi(X)) = E[\varphi(X)]^2 - \left[E[\varphi(X)]\right]^2 = \frac{5}{2}\theta^2 - \left(\frac{3}{2}\theta\right)^2 = \frac{5}{2}\theta^2 - \frac{9}{4}\theta^2 = \frac{1}{4}\theta^2$$

نلاحظ أنّ :

$$Var(Y) = \frac{5}{4}\theta^2 > \frac{1}{4}\theta^2 = Var(\varphi(X))$$

أي أنّ التقدير غير المنحاز $\varphi(X)$ أفضل من التقدير غير المنحاز Y

الجواب الثالث (25 درجة):

نلاحظ أنّ : $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow 1-\frac{\alpha}{2} = 0.975$

وأنّ : $m+n-2 = 10$ ، $n = 5$ ، $m = 7$ وبالتالي فإنّ القيمة المعيارية لتوزيع ستيودنت درجة حريته (10) هي : $t_{1-\frac{\alpha}{2}}(m+n-2) = t_{0.975}(10) = 2.228$.

لنوجد الآن متغيرات العينتين : $S_2^2 = 63.2$ ، $\overline{Y} = 36.8$, $S_1^2 = 61.5$ ، $\overline{X} = 31.1$

وبالتالي فإنّ التباين المكافئ يعطى بالشكل :

$$S_P^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} = \frac{7(61.5)+5(63.2)}{10} = 62.17 \Rightarrow S_P = 7.88$$

وبالتالي فإنّ حدي الثقة المطلوبان هما :

$$L_1 = \left(\overline{X}-\overline{Y}\right) - t_{1-\frac{\alpha}{2}}(m+2-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}} = -5.7 - (2.228)\ (7.88)\ (0.585) = -15.97$$

$$L_2 = \left(\overline{X}-\overline{Y}\right) + t_{1-\frac{\alpha}{2}}(m+n-2)\ S_P\sqrt{\frac{1}{m}+\frac{1}{n}} = -5.7 + (2.228)\ (7.88)\ (0.585) = 4.57$$

الجواب الرابع (20 درجة) :

لدينا $\alpha = 0.10 \Rightarrow 1-\alpha = 0.90 \Rightarrow t_{1-\alpha}(n-1) = t_{0.90}(15) = 1.341$

لنوجد إحصاء الاختبار T بالشكل : $T = \frac{\sqrt{n}\left(\overline{X}-\mu_o\right)}{S} = \frac{4(80-75)}{5} = 4$

نلاحظ أنّ: $T = 4 > 1.341 = t_{1-\alpha}(n-1)$ ومن ثمَّ فإنّنا نرفض الفرضية H_0 عند مستوى الدلالة $\alpha = 0.10$.

انتهت الأجوبة

دمشق في 18 / 8 / 2013 مدرس المقرر: د. سلطان محمد الصلخدي

جامعة دمشق	امتحان مقرر الإحصاء الرياضي	الاســـم : ....
كلية العلوم	لطلاب السنة الثالثة قسم الرياضيات	الدرجــة : مائـــة
قسم الرياضيات	الفصل الأول للعام الدراسي 2013 – 2014	المـدة : ساعتان

السؤال الأول (35 درجة) :

1 - لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي , دالة كثافتها الاحتمالية المشتركة هي $L(\underset{\sim}{x}, \theta)$, وليكن $T = T(X_1, \dots, X_n)$ مقدِّر غير منحاز للوسيط $\varphi(\theta)$, فأثبت أنّ:

$$Var(T) \geq \frac{[\varphi'(\theta)]^2}{E\left(-\frac{\partial^2 \ln L(\underset{\sim}{x},\theta)}{\partial \theta^2}\right)}$$

2 - لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية حجمها n مأخوذة من مجتمع إحصائي يتبع التوزيع البواسوني , قانونه الاحتمالي معرف بالشكل : $f(x,\theta) = e^{-\theta}\frac{\theta^x}{x!}$, حيث $x = 0, 1, 2, 3, \dots$, θ وسيط مجهول , والمطلوب:

1 – أوجد تقدير للوسيط المجهول θ بالطريقة التي تراها مناسبة.

2 – أثبت أنّ المقدِّر الناتج هو مقدِّر غير منحاز ذو تباين أصغري.

السؤال الثاني (25 درجة) :

لنفترض أنّ X_1, X_2, X_3 عينة عشوائية مأخوذة من مجتمع طبيعي معياري , ولنفترض أيضاً أنّ Y_1, Y_2, Y_3 ثلاثة متغيرات عشوائية معرفة بالعلاقات التالية:

$$X_1 = Y_1 \cos Y_2 \sin Y_3 \quad , \quad X_2 = Y_1 \sin Y_2 \sin Y_3 \quad , \quad X_3 = Y_1 \cos Y_3$$

حيث : $0 \leq Y_1 < \infty$, $0 \leq Y_2 \leq 2\pi$, $0 \leq Y_3 \leq \pi$.

والمطلوب: 1 – أوجد دالة الكثافة المشتركة لهذه المتغيرات العشوائية Y_1, Y_2, Y_3 .

2 – أوجد دوال الكثافات الهامشية لكل من Y_1, Y_2, Y_3 .

السؤال الثالث (20 درجة) :

إذا كانت 40 , 30 , 35 , 45 , 35 , 25 , 30 عينة عشوائية مأخوذة من المجتمع الطبيعي $N(\mu_1, \sigma^2)$ وكانت 55 , 50 , 35 , 45 , 65 عينة عشوائية أخرى مستقلة عن الأولى مأخوذة من المجتمع الطبيعي $N(\mu_2, \sigma^2)$ حيث σ^2 مجهولة . أوجد 95 % فترة ثقة للفرق $(\mu_1 - \mu_2)$, وذلك بفرض أنّ القيمة المعيارية هي (2.228) .

السؤال الرابع (20 درجة) :

لتكن X_1, X_2 عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي معرف بدالة الكثافة الآتية:

$$f(x;\theta) = \theta x^{\theta-1} \quad ; 0 < x < 1 \quad ; \quad \theta > 0$$

المطلوب هو استخدم توطئة نيمان - بيرسون في إيجاد منطقة الرفض المثلى في اختبار الفرضية الابتدائية : $H_0 : \theta = 1$ ضد الفرضية البديلة : $H_1 : \theta = 2$.

انتهت الأسئلة

ملاحظة : يسمح باستخدام الآلات الحاسبة

دمشق في 29 / 1 / 2014 مدرس المقرر: د. سلطان محمد الصلخدي

سلم تصحيح مقرر الإحصاء الرياضي

لطلاب السنة الثالثة قسم الرياضيات

الفصل الأول للعام الدراسي 2013 – 2014

الجواب الأول (35 درجة) :

1 – بما أنّ T هو تقدير غير منحاز لـ $\varphi(\theta)$ فإنّ $E(T)=\varphi(\theta)$ لكن:

$$E(T)=\int_{R^n} tL(\underline{x},\theta)d\underline{x}=\varphi(\theta)$$

أي أنّ :

$$\int_{R^n} tL(\underline{x},\theta)d\underline{x}=\varphi(\theta)$$

وباشتقاق العلاقة الأخيرة جزئياً بالنسبة لـ θ نجد :

$$\int_{R^n} t\frac{\partial L(\underline{x},\theta)}{\partial\theta}d\underline{x}=\varphi'(\theta)$$

أي أنّ :

$$\varphi'(\theta)=\int_{R^n} t\frac{\partial L(\underline{x},\theta)}{\partial\theta}\frac{L(\underline{x},\theta)}{L(\underline{x},\theta)}d\underline{x}=\int_{R^n} t\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}L(\underline{x},\theta)d\underline{x}$$

وبالتالي نجد :

$$E\left(T\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)=\varphi'(\theta)$$

وبما أنّ التوقع الرياضي للمشتق الجزئي الأول للدالة $\ln L(\underline{x},\theta)$ يساوي الصفر أي أنّ:

$E\left(\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)=0$ كما وجدنا سابقاً فإنّ:

$$E\left(T\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)=Cov\left(T,\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)$$

ولدينا معامل الارتباط بين متغيرين لا يتجاوز الواحد فإنّ:

$$Cov\left(T,\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)\le\sqrt{VarT}.\sqrt{Var\left(\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)}$$

إذاً :

$$[\varphi'(\theta)]^2\le Var(T).Var\left(\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)$$

وبتقسيم طرفي العلاقة على المقدار الثاني الموجب من الطرف الأيمن نجد:

$$Var(T)\ge\frac{[\varphi'(\theta)]^2}{Var\left(\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)}=\frac{[\varphi'(\theta)]^2}{E\left(\frac{\partial\ln L(\underline{x},\theta)}{\partial\theta}\right)^2}=\frac{[\varphi'(\theta)]^2}{E\left(-\frac{\partial^2\ln L(\underline{x},\theta)}{\partial\theta^2}\right)}$$

2 – – نوجد مقدِّر للوسيط θ بطريقة العزوم :

$$EX=\theta \quad , \ \hat{\theta}=\frac{1}{\mathrm{n}}\sum X_i \quad \Rightarrow \hat{\theta}=\overline{X}$$

لنحسب كل من التوقع الرياضي والتباين فنجد أنّ :

$$E(\overline{X})=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}E(X_i)=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\theta=\theta$$

ونجد أنّ تباينه سيكتب بالشكل التالي :

$$Var(\overline{X})=Var\left[\frac{1}{\mathrm{n}}\sum_{i=1}^{n}E(X_i)\right]=\frac{1}{\mathrm{n}^2}\sum_{i=1}^{n}Var(X_i)=\frac{\theta}{n}$$

لنوجد الآن حد كرامير – راو للمقدِّر غير المنحاز للوسيط θ .

$$\log f(x,\theta)=x\log\theta-\theta-\log(x!)$$

$$\Rightarrow \frac{\partial \log f(x,\theta)}{\partial\theta}=\frac{x}{\theta}-1=\frac{x-\theta}{\theta}$$

ومنه فإنّ :

$$E\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right]=E\left[\left(\frac{X-\theta}{\theta}\right)^2\right]=\frac{1}{\theta^2}E(X-\theta)^2=\frac{1}{\theta^2}\operatorname{Var}(X)=\frac{1}{\theta}$$

وبالتالي سنجد أنّ :

$$B.C.R=\frac{1}{n\left(\frac{1}{\theta}\right)}=\frac{\theta}{n}$$

لنقارن تباين المقدِّر $\overline{X}$ بالحد الأدنى لمتباينة كرامير – راو فنجد أنّهما متساويان وبالتالي فإنّ $\overline{X}$ مقدِّر غير منحاز ذو تباين أصغري .

الجواب الثاني (25 درجة) :

بما أنّ X,Y متغيران عشوائيان مستقلان فإنّ دالة الكثافة المشتركة تعطى بالعلاقة:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3)\overset{\text{من الاستقلال}}{\cong} f_{X_1}(x_1).f_{X_2}(x_2).f_{X_3}(x_3)=\left(\frac{1}{2\pi}\right)^{\frac{3}{2}}e^{-\frac{1}{2}(x_1^2+x_2^2+x_3^2)}$$

$$-\infty<x_1,x_2,x_3<\infty$$

ولدينا : $x_1^2+x_2^2+x_3^2=y_1^2$

بالتعويض في العلاقة نجد :

$$g_{Y_1,Y_2,Y_3}(y_1,y_2,y_2)=f_{X_1,X_2,X_3}(x_1,x_2,x_3)\,|J|_{\substack{x_1=y_1\cos y_2\sin y_3\\ x_2=y_1\sin y_2\sin y_3\\ x_3=y_1\cos y_3}}$$

$$=\left(\frac{1}{2\pi}\right)^{\frac{3}{2}}y_1^2e^{-\frac{1}{2}y_1^2}\sin y_3 \quad ;Y_1\geq 0\,,0\leq Y_2\leq 2\pi\,,0\leq Y_3\leq\pi$$

$$|J| = y_1^2 \sin y_3$$

إيجاد دالة الكثافة الهامشية لكل من Y_1, Y_2, Y_3 .

الجواب الثالث (20 درجة):

لدينا القيمة المعيارية : $t_{1-\frac{\alpha}{2}}(m+n-2) = t_{0.975}(10) = 2.228$.

لنوجد الآن متغيرات العينتين : $\overline{X} = 34.29 \quad , \quad S_1^2 = 45.24 \; , \; \overline{Y} = 50 \quad , \quad S_2^2 = 125$

وبالتالي فإنّ التباين المكافئ يعطى بالشكل :

$$S_P^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} = \frac{6(45.24)+4(125)}{10} = 77.14 \Rightarrow S_P = 8.78$$

وبالتالي فإنّ حدي الثقة المطلوبان هما :

$$L_1 = \left(\overline{X} - \overline{Y}\right) - t_{1-\frac{\alpha}{2}}(m+n-2)\, S_P \sqrt{\frac{1}{m}+\frac{1}{n}} = -15.71 - (2.228)\,(8.78)\,(0.586) = -27.17$$

$$L_2 = \left(\overline{X} - \overline{Y}\right) + t_{1-\frac{\alpha}{2}}(m+n-2)\, S_P \sqrt{\frac{1}{m}+\frac{1}{n}} = -15.71 + (2.228)\,(8.78)\,(0.586) = -4.25$$

الجواب الرابع (20 درجة) :

أنّ منطقة الرفض المثلى هي المجموعة: $C = \left\{(x_1, x_2); \frac{L(\theta')}{L(\theta'')} \le k\right\}$ حيث أنّ k ثابت موجب نحدده من العلاقة: $p\left[(X_1, X_2) \in C \,;\, \text{صحيحة } H_0\right] = 0.05$.

لنوجد الآن $L(\theta') = L(1)$ ثم $L(\theta'') = L(2)$ ثم نأخذ النسبة كما يلي:

$$L(1) = {x_1}^{1-1}.{x_2}^{1-1} = 1$$
$$L(2) = 2{x_1}^{2-1}.2{x_2}^{2-1} = 4x_1x_2$$

$$\frac{L(\theta')}{L(\theta'')} = \frac{L(1)}{L(2)} = \frac{1}{4x_1x_2}$$

وبالتالي فإنّ منطقة الرفض المثلى ما هي إلا مجموعة النقاط :

$$C = \left\{(x_1, x_2): \frac{1}{4x_1x_2} \le k\right\} = \left\{(x_1, x_2): x_1x_2 \ge \frac{1}{4k}\right\}$$

انتهت الأجوبة

مدرس المقرر: د. سلطان محمد الصلخدي

دمشق في 29 / 1 / 2014

جامعة البعث	امتحان مقرر الإحصاء الرياضي	الاســم : ....
كلية العلوم	لطلاب السنة الثالثة قسم الرياضيات	الدرجــة : مائـــة
قسم الرياضيات	الفصل الأول للعام الدراسي 2009 – 2010	المـدة : ساعتان

السؤال الأول (14 درجة) :

ليكن لدينا ثلاثة مجتمعات إحصائية طبيعية ومستقلة : $N(\mu_1,\sigma^2)$, $N(\mu_2,\sigma^2)$, $N(\mu_3,\sigma^2)$ علماً أنّ σ^2 مجهولة . لنفترض أنّه سحبنا عينة عشوائية حجمها n من كل من المجتمعات الإحصائية الثلاثة :

والمطلوب إيجاد فترة ثقة للفرق $(\mu_1 - \mu_2)$ موافق لـ $(1-\alpha)\%$.

السؤال الثاني (16 درجة) :

ليكن لدينا النموذج الخطي البسيط من الشكل : $y_i = \beta_1 + \beta_2 x_i + \epsilon_i \ , \ i = 1,2,\ldots,n$

حيث $x_i \ , i = 1,2,\ldots,n$ مقادير مثبتة ومعروفة ولنفرض أنّ الوسيطين β_1 , β_2 مجهولان وأنّ ϵ_i متغيرات عشوائية مستقلة متوسط كل منها الصفر وتباينه σ^2 , وأنّ : $\beta = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$, $X = \begin{pmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}$

والمطلوب :

1 – استنتج مقدِّري الوسيطين المجهولين β_1 , β_2 .

2 – اكتب عبارة التغاير بين هذين المقدّرين وعبارة التباين لكل منهما .

3 – اكتب عبارة فترة ثقة لـ β_i موافقة لمستوي الثقة $(1-\alpha)\%$.

السؤال الثالث (15 درجة) :

لنفترض أنّ X_1 , X_2 , X_3 عينة عشوائية مأخوذة من المجتمع الإحصائي الموصوف بدالة الكثافة :

$$f(x) = e^{-x} \quad , \quad 0 < x < \infty$$

ولنفترض أيضاً أنّ Y_1 , Y_2 , Y_3 ثلاثة متغيرات عشوائية أخرى معرفة بالعلاقات التالية :

$$Y_1 = \frac{X_1}{X_1+X_2} \quad , \quad Y_2 = \frac{X_1+X_2}{X_1+X_2+X_3} \quad , \quad Y_3 = X_1+X_2+X_3$$

والمطلوب إيجاد دالة الكثافة الهامشية لكل من هذه المتغيرات العشوائية Y_1 , Y_2 , Y_3 , ثم أثبت أنّها مستقلة .

السؤال الرابع (20 درجة) :

ليكن X, Y متغيران عشوائيان دالة كثافتهما الاحتمالية المشتركة :

$$f_{X,Y}(x,y) = \frac{2}{\vartheta^2} e^{-\frac{x+y}{\vartheta}} \qquad ; 0 < x < y < \infty \ , \vartheta > 0$$

والمطلوب :

1 – أوجد التوقع الرياضي للمتغير العشوائي Y ثم أوجد تباينه .

2 – أوجد التوقع الرياضي الشرطي للمتغير العشوائي Y علماً $X = x$ ولنرمز له بالرمز $\varphi(x)$.

3 – أحسب تباين المقدِّر $\varphi(x)$ ثم بيّن أي المقدِّرين أفضل Y أم $\varphi(x)$.

السؤال الخامس (15 درجة) :

لتكن X_1 , X_2 عينة عشوائية حجمها $n = 2$ مأخوذة من مجتمع إحصائي معرف بدالة الكثافة الاحتمالية التالية :

$$f(x,\vartheta) = \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}} \ , \ x > 0$$

والمطلوب إيجاد منطقة الرفض المثلى في اختبار الفرضية الابتدائية : $H_0 : \vartheta = 3$ ضد الفرضية البديلة : $H_1 : \vartheta = 5$

وذلك عند مستوى الأهمية $\alpha = 0.05$.

انتهت الأجوبة

دمشق في 2010 / 1 / 19 — مدرس المقرر: د. سلطان محمد الصلخدي

سلم تصحيح مقرر الإحصاء الرياضي

لطلاب السنة الثالثة قسم الرياضيات

الفصل الأول للعام الدراسي 2009 – 2010

الجواب الأول (14 درجة):

نبحث عن الكمية المحورية المناسبة :

$$\left.\begin{array}{l} X\sim N(\mu_1,\sigma^2) \\ Y\sim N(\mu_2,\sigma^2) \\ Z\sim N(\mu_3,\sigma^2) \end{array}\right\} \Rightarrow \left.\begin{array}{l} \bar{X}-\mu_1\sim N\left(0,\frac{\sigma^2}{n}\right) \\ \bar{Y}-\mu_2\sim N\left(0,\frac{\sigma^2}{n}\right) \\ \bar{Z}-\mu_3\sim N\left(0,\frac{\sigma^2}{n}\right) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \bar{X}-\bar{Y}-(\mu_1-\mu_2)\sim N\left(0,\frac{2\sigma^2}{n}\right) \\ \bar{X}-\bar{Z}-(\mu_1-\mu_3)\sim N\left(0,\frac{2\sigma^2}{n}\right) \\ \bar{Y}-\bar{Z}-(\mu_2-\mu_3)\sim N\left(0,\frac{2\sigma^2}{n}\right) \end{array}\right.$$

و هذه العلاقة تؤدي إلى :

$$\bar{X}-\bar{Y}-(\mu_1-\mu_2)\sim N\left(0,\frac{2\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{2}{n}}}\sim N(0,1)$$

ولدينا :

$$\left.\begin{array}{l} s_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \\ s_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 \\ s_3^2=\frac{1}{n-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2 \end{array}\right\} \Rightarrow s^2=\frac{s_1^2+s_2^2+s_3^2}{3} \Rightarrow 3s^2=s_1^2+s_2^2+s_3^2\}$$

$$\Rightarrow \frac{3(n-1)s^2}{\sigma^2}=\underbrace{\frac{(n-1)s_1^2}{\sigma^2}}_{\chi^2_{n-1}}+\underbrace{\frac{(n-1)s_2^2}{\sigma^2}}_{\chi^2_{n-1}}+\underbrace{\frac{(n-1)s_3^2}{\sigma^2}}_{\chi^2_{n-1}}\sim\chi^2_{3n-3}$$

$$\Rightarrow \frac{3(n-1)s^2}{\sigma^2}\sim\chi^2_{3n-3}$$

واعتمادا على نظرية سابقة نجد أنّ الكمية :

$$T=\frac{Z\sim N(0,1)}{\sqrt{\frac{Y\sim\chi^2_n}{n}}}=\frac{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{2}{n}}}}{\sqrt{\frac{\frac{3(n-1)s^2}{\sigma^2}}{3n-3}}}=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s\sqrt{\frac{2}{n}}}\sim t_{3n-3}$$

أنّ هذه الكمية تصلح لأنّ تكون كمية كحورية لأنّها تحوي $\mu_1-\mu_2$ مجهولة توزيعها هو توزيع ستيودنت.
نحصر هذه الكمية بين قيمتين معياريتين ونجدهما من جدول توزيع ستيودنت درجة حريته $(3n-3)$ ثم عزل $\mu_1-\mu_2$ فنحصل على فترة الثقة المطلوبة :

$$[L_1,L_2]=\left[\bar{X}-\bar{Y}-t_{1-\alpha,3n-3}\ s\sqrt{\frac{2}{n}}\ ,\ \bar{X}-\bar{Y}+t_{1-\alpha,3n-3}\ s\sqrt{\frac{2}{n}}\right]$$

الجواب الثاني (16 درجة):

لدينا :

$$\beta=\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}\ ,\ X=\begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{bmatrix}$$

ومن السهل رؤية أنّ :

$$\left.\begin{array}{c} S=X^{\prime}X=\begin{pmatrix} n & \sum x_i\\ \sum x_i & \sum x_i{}^2\end{pmatrix}\\ X^{\prime}Y=\begin{pmatrix}\sum y_i\\ \sum x_i y_i\end{pmatrix}\end{array}\right\}\Longrightarrow S^{-1}=\frac{1}{n\sum(x_i-\bar{x})^2}\begin{pmatrix}\sum x_i{}^2 & -\sum x_i\\ -\sum x_i & n\end{pmatrix}$$

و هكذا نجد أنّ :

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=S^{-1}X^{\prime}Y$$

$$=\frac{1}{n\sum(x_i-\bar{x})^2}\begin{pmatrix}\sum x_i{}^2\sum y_i-\sum x_i\sum x_i y_i\\ -\sum x_i\sum y_i+n\sum y_i x_i\end{pmatrix}_{2\times 1}$$

أي أنّ :

$$\hat{\beta}_2=\frac{\sum(x_i-\bar{x})(y_i-y)}{\sum(x_i-\bar{x})^2}\quad ,\ \hat{\beta}_1=\hat{y}-\hat{\beta}_2\bar{x}$$

و β_1 هو ميل خط الانحدار $E(y/x)=\beta_1+\beta_2 x$ و β_1 هو قيمة $E(Y)$ عندما $x=0$. وبما أنّ $Cov(\hat{\beta})=S^{-1}\sigma^2$ فإنّنا نجد :

$$Cov(\hat{\beta})=S^{-1}\sigma^2\Longrightarrow Cov(\hat{\beta}_1,\hat{\beta}_2)=-\frac{\sigma^2\sum x_i}{n\sum(x_i-\bar{x})^2}$$

$$Var(\hat{\beta}_1)=\frac{\sigma^2\sum x_i{}^2}{\sum(x_i-\bar{x})^2}\quad ,\quad Var(\hat{\beta}_2)=\frac{\sigma^2}{n\sum(x_i-\bar{x})^2}$$

$$\left[\hat{\beta}_i-t_{1-\frac{\alpha}{2}}\ \hat{\sigma}\sqrt{C_{ii}}\ ,\ \hat{\beta}_i+t_{1-\frac{\alpha}{2}}\ \hat{\sigma}\sqrt{C_{ii}}\right]$$

الجواب الثالث (15 درجة):

لنوجد أولا دالة الكثافة الاحتمالية المشتركة للمتغيرات X_1,X_2,X_3 :

$$f(x_1,x_2,x_3)=\prod_{i=1}^{3}e^{-x_i}=e^{-(x_1+x_2+x_3)}=e^{-y_3}$$

ثم نوجد دالة الكثافة الاحتمالية المشتركة للمتغيرات Y_1,Y_2,Y_3 :

$$g(y_1,y_2,y_3)=f(x_1,x_2,x_3)|J|\ \dots(1)$$

لدينا :

$$\left.\begin{array}{c} y_1=\frac{x_1}{x_1+x_2}\\ y_2=\frac{x_1+x_2}{x_1+x_2+x_3}\\ y_3=x_1+x_2+x_3\end{array}\right\}\Longrightarrow\left\{\begin{array}{c} x_1=y_1y_2y_3\\ x_2=(1-y_1)y_2y_3\\ x_3=y_3(1-y_2)\end{array}\right.$$

وبالتالي فإنّ معين اليعقوبي J يعطى كما يلي:

$$J=\begin{vmatrix}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3}\\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3}\\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3}\end{vmatrix}=\begin{vmatrix} y_2y_3 & y_1y_3 & y_2y_1\\ -y_2y_3 & (1-y_1)y_3 & (1-y_1)y_2\\ 0 & -y_3 & 1-y_2\end{vmatrix}=-y_2y_3^2$$

نعوض في (1) فنجد أنّ :

$$g(y_1,y_2,y_3)=e^{-y_3}y_2y_3^2=y_2y_3^2e^{-y_3}$$

علماً بأنّ: $\quad 0<y_1<1 \qquad 0<y_2<1 \qquad 0<y_3<\infty$

$$g(y_1)=\int_0^1\int_0^\infty y_2y_3^2e^{-y_3}\,dy_2\,dy_3=1 \;\;,\; 0<y_1<1$$

$$g(y_2)=\int_0^1\int_0^\infty y_2y_3^2e^{-y_3}\,dy_1\,dy_3=2y_2 \;,\;\; 0<y_2<1$$

$$g(y_3)=\int_0^1\int_0^1 y_2y_3^2e^{-y_3}\,dy_1\,dy_2=\frac{y_3^2e^{-y_3}}{2} \;\;,0<y_3<\infty$$

$$g(y_1,y_2,y_3)=e^{-y_3}y_2y_3^2=y_2y_3^2e^{-y_3}=g(y_1)g(y_2)g(y_3)$$

وبالتالي فإنّ المتغيرات Y_1,Y_2,Y_3 مستقلة وبذلك يتم المطلوب .

الجواب الرابع (20 درجة):

$$f_X(x)=\int_{y=x}^{y=\infty}f_{X,Y}(x,y)dy=\int_{y=x}^{y=\infty}\frac{2}{\theta^2}e^{-\frac{x+y}{\theta}}\,dy$$

$$=\frac{2}{\theta}e^{-\frac{x}{\theta}}\int_{y=x}^{y=\infty}\frac{1}{\theta}e^{-\frac{y}{\theta}}\,dy=\frac{2}{\theta}e^{-\frac{x}{\theta}}\left[-e^{-\frac{y}{\theta}}\right]_{y=x}^{y=\infty}$$

$$=\frac{2}{\theta}e^{-\frac{x}{\theta}}\left[\,e^{-\frac{x}{\theta}}\right]=\frac{2}{\theta}e^{-\frac{2x}{\theta}} \qquad ;\theta>0,x>0$$

أي أنّ X أسي وسيطه $\lambda=\frac{2}{\theta}$.

$$f_y(y)=\int_{x=0}^{x=y}f_{X,Y}(x,y)dx=\int_{x=0}^{x=y}\frac{2}{\theta^2}e^{-\frac{x+y}{\theta}}\,dx$$

$$=\frac{2}{\theta}e^{-\frac{y}{\theta}}\int_{x=0}^{x=y}\frac{1}{\theta}e^{-\frac{x}{\theta}}\,dy=\frac{2}{\theta}e^{-\frac{y}{\theta}}\left[-e^{-\frac{x}{\theta}}\right]_{x=0}^{x=y}$$

$$=\frac{2}{\theta}e^{-\frac{y}{\theta}}\left[1-e^{-\frac{y}{\theta}}\right]=\frac{2}{\theta}e^{-\frac{y}{\theta}}-\frac{2}{\theta}e^{-\frac{2y}{\theta}} \qquad ;\theta>0,y>0$$

لنوجد التوقع الرياضي:

$$E(Y)=\int_0^\infty yf_y(y)dy=\int_0^\infty y\left[\frac{2}{\theta}e^{-\frac{y}{\theta}}-\frac{2}{\theta}e^{-\frac{2y}{\theta}}\right]dy=\frac{2}{\theta}\underbrace{\int_0^\infty y\,e^{-\frac{y}{\theta}}dy}_{I_1}-\frac{2}{\theta}\underbrace{\int_0^\infty y\,e^{-\frac{2y}{\theta}}dy}_{I_2}$$

$$I_1=\int_0^\infty y\,e^{-\frac{y}{\theta}}dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{1}{\theta}y=\mathfrak{z}\Rightarrow dy=\theta d\mathfrak{z}$$

وبالتالي فإنّ :

$$I_1=\theta^2\int_0^\infty \mathfrak{z}\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^2\int_0^\infty \mathfrak{z}^{2-1}\,e^{-\mathfrak{z}}d\mathfrak{z}=\theta^2\Gamma(2)=\theta^2$$

حساب التكامل الثاني:

$$I_2 = \int_0^\infty y\, e^{-\frac{2y}{\theta}} dy$$

لحساب هذا التكامل نفرض أنّ :

$$\frac{2}{\theta} y = t \Rightarrow dy = \frac{\theta}{2} dt$$

وبالتالي فإنّ :

$$I_2 = \frac{\theta^2}{4} \int_0^\infty t\, e^{-t} dt = \frac{\theta^2}{4} \int_0^\infty t^{2-1}\, e^{-t} dt = \frac{\theta^2}{4} \Gamma(2) = \frac{\theta^2}{4}$$

$$E(y) = 2.\frac{1}{\theta}.\theta^2 - 2.\frac{1}{\theta}.\theta^4 = 2\theta - \frac{1}{2}\theta = \frac{3}{2}\theta \quad Var(y) = \underbrace{Ey^2}_{?} - [E(y)]^2$$

$$E(y^2) = \int_0^\infty y^2 f_y(y) dy = \int_0^\infty y^2 \left[\frac{2}{\theta} e^{-\frac{y}{\theta}} - \frac{2}{\theta} e^{-\frac{2y}{\theta}}\right] dy$$

$$= \frac{2}{\theta} \underbrace{\int_0^\infty y^2\, e^{-\frac{y}{\theta}} dy}_{I_1} - \frac{2}{\theta} \underbrace{\int_0^\infty y^2\, e^{-\frac{2y}{\theta}} dy}_{I_2}$$

$$I_1 = \int_0^\infty y^2\, e^{-\frac{y}{\theta}} dy$$

$$\frac{1}{\theta} y = z \Rightarrow dy = \theta dz \Rightarrow I_1 = \theta^3 \int_0^\infty z^2\, e^{-z} dz = \theta^3 \int_0^\infty z^{3-1}\, e^{-z} dz = \theta^3 \Gamma(3) = 2\theta^3$$

$$I_2 = \int_0^\infty y^2\, e^{-\frac{2y}{\theta}} dy$$

$$\frac{2}{\theta} y = t \Rightarrow dy = \frac{\theta}{2} dt \Rightarrow I_2 = \frac{\theta^3}{8} \int_0^\infty t^2\, e^{-t} dt = \frac{\theta^3}{8} \int_0^\infty t^{3-1}\, e^{-t} dt = \frac{\theta^3}{8} \Gamma(3) = \frac{\theta^3}{4}$$

$$E(y^2) = \frac{2}{\theta}\, 2\theta^3 - \frac{2}{\theta} \frac{\theta^3}{4} = 4\theta^2 - \frac{1}{2}\theta^2 = \frac{7}{2}\theta^2$$

$$Var(y) = Ey^2 - [E(y)]^2 = \frac{7}{2}\theta^2 - \left(\frac{3}{2}\theta^2\right)^2 = \frac{5}{4}\theta^2$$

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{2}{\theta^2} e^{-\frac{x+y}{\theta}}}{\frac{2}{\theta} e^{-\frac{2x}{\theta}}} = \frac{1}{\theta}\, e^{-\frac{1}{\theta}(y-x)} ; \begin{cases} 0 < x < y < \infty \\ \quad 0 < x \end{cases}$$

$$E(y/x) = \int_{y=x}^{y=\infty} y\, f_{Y/X}(y/x) dy = \int_{y=x}^{y=\infty} y \left[\frac{1}{\theta}\, e^{-\frac{1}{\theta}(y-x)}\right] dy$$

$$= \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{y=x}^{y=\infty} y\, e^{-\frac{1}{\theta} y} dy = \frac{1}{\theta}\, e^{\frac{x}{\theta}} \int_{z=\frac{x}{\theta}}^{z=\infty} \theta^2\, z\, e^{-z} dz$$

$$= \theta\, e^{\frac{x}{\theta}} \int_{z=\frac{x}{\theta}}^{z=\infty} z\, e^{-z} dz$$

نكامل بالتجزئة لذلك نفرض :

$$\left.\begin{aligned} u = z \Rightarrow du = dz \\ dv = e^{-z} dz \Rightarrow v = -e^{-z} \end{aligned}\right\} \Rightarrow E(y/x) = \theta\, e^{\frac{x}{\theta}} \left[-z\, e^{-z}\Big|_{z=\frac{x}{\theta}}^{z=\infty} + \int_{z=\frac{x}{\theta}}^{z=\infty} e^{-z} dz\right]$$

$$= \theta\, e^{\frac{x}{\theta}} \left[\frac{x}{\theta}\, e^{-\frac{x}{\theta}} + e^{-\frac{x}{\theta}}\right] = x + \theta = \varphi(x)$$

$$Var(\varphi(x)) = \underbrace{E[\varphi(x)]^2}_{?} - \underbrace{(E[\varphi(x)])^2}_{?}$$

$$E[\varphi(x)] = E[x+\theta] = E[x] + \theta = \frac{\theta}{2} + \theta = \frac{3}{2}\theta$$

$$E[\varphi(x)]^2 = E[x+\theta]^2 = E[x^2 + 2\theta x + \theta^2]$$

$$= E[x^2] + 2\theta E[x] + \theta^2$$

$$= E[x^2] + 2\theta\left[\frac{\theta}{2}\right] + \theta^2$$

$$= \frac{\theta^2}{2} + \theta^2 + \theta^2 = \frac{5}{2}\theta^2$$

حيث لدينا X أسي وسيطه $\lambda = \frac{2}{\theta}$.

$$\Rightarrow Var(X) = E[X^2] - [E(X)]^2 \Rightarrow$$

$$E[X^2] = Var(X) + [E(X)]^2$$

$$= \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2 = \frac{\theta^2}{4} + \frac{\theta^2}{4} = \frac{\theta^2}{2} \Rightarrow$$

$$Var(\varphi(X)) = E[\varphi(X)]^2 - \left[E[\varphi(X)]\right]^2 = \frac{5}{2}\theta^2 - \left(\frac{3}{2}\theta\right)^2 = \frac{5}{2}\theta^2 - \frac{9}{4}\theta^2 = \frac{1}{4}\theta^2$$

نلاحظ أنّ :

$$Var(Y) = \frac{5}{4}\theta^2 > \frac{1}{4}\theta^2 = Var(\varphi(X))$$

أي أنّ التقدير غير المنحاز $\varphi(X)$ أفضل من التقدير غير المنحاز Y

<u>الجواب الخامس (15 درجة):</u>

أنّ منطقة الرفض المثلى هي المجموعة: $C = \left\{(x_1, x_2); \frac{L(\theta')}{L(\theta'')} \le k\right\}$ حيث أنّ k ثابت موجب نحدده من العلاقة : $p\left[(X_1, X_2, \dots, X_n) \in C \text{ ; صحيحة } H_0\right] = \alpha = 0.05$

لنوجد الآن $L(\theta') = L(3)$ ثم $L(\theta'') = L(5)$ ثم نأخذ النسبة كما يلي:

$$L(3) = \frac{1}{3}e^{-\frac{x_1}{3}}.\frac{1}{3}e^{-\frac{x_2}{3}} = \frac{1}{9}e^{-\left(\frac{x_1+x_2}{3}\right)}$$

$$L(5) = \frac{1}{5}e^{-\frac{x_1}{5}}.\frac{1}{5}e^{-\frac{x_2}{5}} = \frac{1}{25}e^{-\left(\frac{x_1+x_2}{5}\right)}$$

لنأخذ النسبة :

$$\frac{L(\theta')}{L(\theta'')} = \frac{L(3)}{L(5)} = \frac{25}{9}e^{-\left(\frac{x_1+x_2}{3}\right)+\left(\frac{x_1+x_2}{5}\right)} = \frac{25}{9}e^{-\frac{2}{15}(x_1+x_2)}$$

وبالتالي فإنّ منطقة الرفض المثلى ما هي إلا مجموعة النقاط :

$$C = \left\{(x_1, x_2): \frac{25}{9}e^{-\frac{2}{15}(x_1+x_2)} \le k\right\} = \left\{(x_1, x_2): x_1 + x_2 \ge -\frac{15}{2}\mathrm{Ln}\left(\frac{9k}{25}\right)\right\}$$

<u>ملاحظة</u> : إذا ارتكب الطالب خطأً حسابياً فتحذف له علامة تقديرية حسب تأثير هذا الخطأ على السؤال .

انتهت الأجوبة

حمص في 19 / 1 / 2010 — **مدرس المقرر: د. سلطان محمد الصلخدي**

مثال: لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية مأخوذة من مجتمع إحصائي دالة كثافته معرفة بالشكل:

$$f(x,\theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}} \quad ; \quad x \geq 0 \ ,\left(\text{وسيط } \theta > 0\right)$$

والمطلوب:

1 – أوجد مقدِّر للوسيط θ .

2 – أثبت أنّ هذا المقدِّر $\hat{\theta}$ هو مقدِّر غير منحاز ومتسق وكاف وذو تشتت أصغري.

الحل:

1– لنقدر الوسيط θ باستخدام طريقة الاحتمالية العظمى:

$$L\left(\underline{x},\theta\right) = \prod_{i=1}^{n} f(x_i,\theta) = \frac{2^n \prod_{i=1}^{n} x_i}{\theta^n} e^{-\sum_{i=1}^{n} \frac{{x_i}^2}{\theta}}$$

نأخذ لوغاريتم دالة الاحتمالية:

$$\ln L(\underline{x},\theta) = n\ln(2) - n\ln(\theta) - \sum_{i=1}^{n} \frac{{x_i}^2}{\theta} + \sum_{i=1}^{n} \ln(x_i) \ \dots (*)$$

نشتق جزئياً بالنسبة للوسيط θ فنجد:

$$\frac{\partial \ln L(\underline{x},\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{\mathrm{i}=1}^{\mathrm{n}} {x_i}^2$$

نستبدل كل θ بـ $\hat{\theta}$ ثم نجعل هذا المقدار مساوياً للصفر ونحل المعادلة:

$$-\frac{n}{\hat{\theta}} + \frac{1}{\left(\hat{\theta}\right)^2} \sum_{i=1}^{n} {x_i}^2 = 0 \Longrightarrow \frac{n}{\hat{\theta}} = \frac{1}{\left(\hat{\theta}\right)^2} \sum_{i=1}^{n} {x_i}^2$$

$$\Longrightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} {x_i}^2$$

2– نفرض أنّ $t = \sum_{i=1}^{n} {x_i}^2$ ثم نبدل في $(*)$:

$$\ln L(\underline{x},\theta) = n\ln(2) - n\ln(\theta) - \frac{t}{\theta} + \sum_{i=1}^{n} ln(x_i) = \theta_2 + \theta_1 t + \ln L_1$$

حيث $\theta_1 = -\frac{n}{\theta}$, $\theta_2 = -n\ln\theta$, $\ln(L_1) = n\ln 2 + \sum_{i=1}^{n} {x_i}^2$

ومنه فإنّ:

$$L = L_1 e^{\theta_1 t + \theta_2}$$

نوجد التوقع الرياضي لـ t , لدينا:

$$E\ (T) = -\frac{d\theta_2}{d\theta_1} = -\frac{\frac{d\theta_2}{d\theta}}{\frac{d\theta_1}{d\theta}} = -\frac{-\frac{n}{\theta}}{\frac{n}{\theta^2}} = \theta$$

فالتقدير T هو تقدير غير منحاز.

$$Var\ (t) = -\frac{d^2\theta_2}{d\theta_1^2} = \left(\frac{d\theta_1}{d\theta}\right)^{-1} . \frac{d}{d\theta}\left(\frac{d\theta_2}{d\theta_1}\right) = \frac{\theta^2}{n} L = \frac{\theta^2}{n}$$

إذاً التقدير $\hat{\theta}$ هو تقدير غير منحاز تباينه يبلغ حد المعلومات وهو متسق لأنّه غير منحاز وتباينه يسعى إلى الصفر عندما $n \to \infty$, وهو كافٍ لأنّ دالة الكثافة الاحتمالية كتبت على الشكل: $L = L_1 e^{\theta_1 t + \theta_2}$, وبالتالي فهو ذو تشتت أصغري.

لتكن لدينا $X_1, \dots, \mathrm{X_n}$عينة عشوائية مأخوذة من مجتمع إحصائي دالة كثافته معرفة بالشكل التالي:

$$f(x,\theta) = \frac{\alpha^{\theta}}{\Gamma(\theta)} e^{-\alpha x} x^{\theta-1} \quad , \quad \alpha, \theta > 0 \quad ; \quad x \geq 0$$

دالة الكثافة لهذه المتغيرات ما هي إلا:

$$L(\underline{x}, \alpha, \theta) = \frac{\alpha^{n\theta}}{[\Gamma(\theta)]^n} e^{-\alpha \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i^{\theta-1}$$

نفرض أنّ θ معلومة ولنكتب : $L = L_1 e^{-yn\theta x + n\theta \ln \alpha}$ حيث $y = \frac{\sum_{i=1}^{n} x_i}{n\theta}$

وهكذا يكون $L_1 = \frac{\prod_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i}^{\theta-1}}{[\Gamma(\theta)]^n}$

ويكون بالتالي L_1 مستقلاً عن α والحد الأدنى للتشتت يدركه التقدير y للكمية : $\frac{1}{\alpha} = \frac{\frac{d}{d\alpha}(n\theta \ln \alpha)}{\frac{d}{d\alpha}(-n\theta\alpha)}$

ولدينا: $\quad var\,(Y) = \frac{1}{\frac{d}{d\alpha}(-n\theta\alpha)} \frac{d}{d\alpha} \frac{1}{\alpha} = \frac{1}{n\theta\alpha^2}$

نفرض أنّ α معلومة ونكتب $L = L_1 e^{zn(\theta-1) - n \ln \Gamma(\theta) + n\theta \ln \alpha}$

حيث أنّ: $\mathrm{z} = \frac{1}{\mathrm{n}} \ln \prod_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i}$ و $L_1 = e^{-\alpha \sum \mathrm{x_i}}$ هو مستقل عن θ

والتقدير الذي يدرك تباينه الحد الأدنى هو التقدير Z للكمية

$$\frac{\frac{d}{d\theta}(n \ln \Gamma(\theta) - n\theta \ln \alpha)}{\frac{d}{d\theta}(n(\theta-1))} = \frac{d}{d\theta} \ln \Gamma(\theta) - \ln \alpha$$

ولدينا $var\, Z = \frac{1}{n} \frac{d^2}{d\theta^2} \ln \Gamma(\theta)$