STATISTICAL MODELS PRODUCED FROM FISHER INFORMATION FUNCTION

Prof. Dr. Ahmed Al – Aloosy
Azzam A.Tawfiq
Baghdad College of Economic Sciences University

Abstract

Statistics the science of extracting information from data appears the most natural field of application of information theoretic methods in statistics. the Fisher information I (θ) is the variance of score . it is named in honor of its inventor the statistician R. A .Fisher .the fisher information is the amount of information that an observable random variable X carries about unobservable parameter θ upon which the likelihood function of $X, L(\theta) = F(X;\theta)$,depends.

The likelihood function is the joint probability of the data , the Xs , conditional on the value of $\theta,$ as a function of $\theta.$ Since the expectation of the score is zero , the variance is simply the second moment of the likelihood function with respect to $\theta.$ Hence the Fisher information can be written

$I(\theta)$ = E { $[\vartheta/\vartheta\theta \ln f\ (x,\theta)]^2|\theta$ } Which implies $0 \leq I(\theta)<\infty$.

We are discussing the regular estimation case when:

1- The range of the random variable X does not depend upon the unknown parameter θ .
   i.e $a\leq x \leq b$ where a and b are constants .
2- Differentiation with respect to θ can be carried out under the integral $\int f\ (x,\theta)dx$ where the limits from c (θ) $to\ d$ (θ) depend on θ .
   Both univariate and multivariate parameters using fisher information matrix .

Key words: maximum likelihood , Leibnitz rule , fisher information matrix.

الملخص

إن نظرية المعلوماتية تعني إستخلاص اعظم درجة من المعلومات الماخوذة من العينة حول معالم المجتمع .
العالم فشر هو الذي ابتكر طريقة تقدير معالم المجتمع باستخدام طريقة الامكان الاعظم ، ويمكن تعميم ذلك الى اكثر من عينة لتقدير معالم المجتمع بواسطة ايجاد التغاير المشترك واستخراج المصفوفة المناسبة.
ان معادلة فشر للمعلوماتية يمكن كتابتها بالصيغة التالية

$I(\theta)$ = E { $[\vartheta/\vartheta\theta \ln f\ (x,\theta)]^2 \mid \theta$ } where $0 \leq I(\theta)<\infty$

لقد تم استخدام نظرية ليبنتز في حالة التكامل عندما تكون النهايات دوال للمتغير.

Fisher information function and information matrix:

Case (1): one parameter

Let f (x, θ) be the p.d.f or a r .v. x , may be discrete or continuous, we assume that f (x,θ) depends upon a continuous parameter θ.
Let Ω denote the parameter space , we also assume that the range of the r.v . x does not depend upon the unknown parameter θ .

The regular estimation case :

To have a regular estimation case when :

a- The range of the r.v.x does not depend upon the unknown parameter θ
i.e a ≤ x ≤ b

b- Differentiation w.r.t. θ can be carried out under the integral sign .i.e$\int_c^d f(x,\theta)dx$ or we can differentiate the integral

$$\frac{d}{d\theta}\int_c^d f(x,\theta)dx = \int_c^d \{\frac{d}{d\theta} f(x,\theta)\}dx$$

(In condition (b) if the limits c and b are functions of θ , say c (θ) and d (θ), then

$$\frac{d}{d\theta}\int_{c(\theta)}^{d(\theta)} f(x,\theta)dx = \int_{c(\theta)}^{d(\theta)} \frac{d}{d\theta} f(x,\theta)dx + f\{d(\theta),\theta\}d'(\theta) - f\{c(\theta),\theta\}d\, c'(\theta)$$

Deferent rule for differentiation under integral sign).

Since { $\int_a^b f(x,\theta)dx = 1$}for any $\theta \in \Omega$

Differentiate both sides w. r.t. θ we set:

$$\frac{d}{d\theta}\int_a^b f(x,\theta)dx = \int_a^b \frac{d}{d\theta} f(x,\theta)dx = 0 \ldots\ldots\ldots\ldots\ldots (1)$$

Or$\int_a^b \frac{d\log f(x,\theta)}{d\theta}\ f(x,\theta)dx = 0$ ………………………….(2)

OrE$\frac{d}{d\theta}\ log f(x,\theta)$)=0 …………………………….(3)

Differentiate (2) w.r.t θ under the integral sign .

$$\int_a^b [(\frac{d\,2}{d\theta^2}\ log\, f(x,\theta))\, f(x,\theta) + (\frac{d}{d\theta}\ log\, f(x,\theta)\frac{df}{d\theta}(x,\theta)]\, dx = 0$$

$$\therefore \int_a^b (\frac{d}{d\,\theta^2}\ log\, f\ (x,\theta)) f(x,0)dx + \int_a^b (\frac{d}{d\theta}\ log\, f\ (x,\theta))^2\ f(x,\theta)\ dx = 0$$

$\therefore$ E $\left(\frac{d2 \ log \ f(x,\theta)}{d \ \theta^2} \right)$ = E $\left(\frac{d}{d\theta} \ log \ f \ (x,\theta) \right)^2$=I($\theta$) ……………………..(4)

Where I (θ) $is\ called\ fishers\ , information\ function$.

$\boldsymbol{Note}$: $that\ from$ (3) and(4)
Var $\left(\frac{d}{d\theta} \ log \ f \ (x,\theta)\right) = I \ (\theta) \geq 0$

Likelihood function and I(θ) :

Let x_1 , x_2 , … x_n be n independent observations on the r.v. X , then the function
L= $\prod_{i=1}^{n} f \ (xi,\theta)$

Is called the likelihood function of the given sample

$$\text{Log L} = \sum_{i=1}^{n} \log f \ (x_{\ i} \ , \theta \ )$$

$$\frac{d}{d \ \theta}\log L = \sum_{i=1}^{n} \log f \ (x_{\ i} \ , \theta \ )$$

$\therefore$ E $\left(\frac{d}{d \ \theta}\log L \right)$= $\sum_{i=1}^{n} E \left\{ \frac{d}{d \ \theta}\log f \ (x,\theta) \right\} = 0$ …………….. (5)

Also :

$\frac{d2 \log l}{d \ \theta^{\ 2}} = \sum_{i=1}^{n} \frac{d2}{d\theta^2}\log$ f(x_i , θ)
$\therefore$ E $\left(-\frac{d2 \log l}{d\theta^{\ 2}} \right)$= $\sum_{i=1}^{n} E \left\{ \frac{d2}{d \ \theta^2}\log f \ (x_i,\theta) \right\} = n \ I(\ \theta)$ ……………….(6)

From (5)

$\left(\frac{d}{d\theta} \log l \right)^2 = \sum_{i}^{n} \sum_{i}^{n} \frac{d}{d\theta}\log f \ (x_i,\theta) \frac{d}{d\theta}\log f \ (x_i,\theta)$
$\therefore E\left(\frac{d}{d\theta}\log l\right)^2$ = E $\left\{ \frac{d}{d\theta}\log f(x_i,\theta) . \frac{d}{d\theta}\log f \ (x_i,\theta) \right\}$

Since x_1, x_2 ,………………….., x_n are independent , so are the function

$\frac{d}{d\theta}\log$ f (x_i ,θ),……………., $\frac{d}{d\theta}$ log f(x_n,θ).This gives
E $\left(\frac{d}{d\theta} \log l \right)^2 = \sum_{i=1}^{n} E \left\{ \frac{d}{d\theta}\log f \ (x_i,\theta \ ) \right\}^2$ = n I(θ)…………. (7)

From (6) and (7)$we\ observe$

E $\left(\frac{d}{d\theta}\log l \right)^2 = E \left(- \frac{d2}{d\theta^2}\log l \right)$= n I(θ) ……….. (8)

Where n I (θ) is called fisher's amount information contained in the sample about the parameter θ.

Example :
Let x havebinomial dist.
X=1 iftrial givessuccess
X=0 otherwise
If p= p(x=1) ,and q= p(x=0) (where p+q=1)

Then x has the p .d. f .

$f(x, p) = p(X=x) = p^x (1-p)^{1-x}$

$\therefore \log f(x, p) = x \log p + (1-x) \log(1-p)$

$$\frac{d \log f(x,p)}{dp} = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\therefore E\left(\frac{d \log f(x,p)}{dp}\right) = \frac{E(x)}{p} - \frac{E(1-X)}{1-P} = \frac{P}{P} - \frac{1-P}{1-P} = 0$$

$$\frac{d2}{dp^2} \log f(x,p) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

$$I(P) = E\left(-\frac{d2}{dp2} \log f(x,p)\right) = \frac{E(X)}{P^2} + \frac{1-E(X)}{(1-P)^2} = 0$$

$$= \frac{P}{P^2} + \frac{1-P}{(1-P)2} = \frac{1}{P} + \frac{1}{1-P} = \frac{1}{P(1-P)}$$

Alternativelyif $x_1, x_2, \ldots, x_n$ are independent observation x
i. e p (x=1)=p ifi [th]trial is success
= 0 otherwise
And if $Z = x_1 + x_2 + \ldots + x_n$
Then the p .d. forthe r. v. Z in the same likelihood function
Function of the givensamplei.e

$P(Z=z) = (z^n)\, p^z (1-p)^{n-z}$

We get

$n\,I(p) = \frac{n}{p(1-p)}$.

Example (2) : consider the gamma dist function having the p . d .f .

$F(x, \theta, l) = \frac{1}{\theta^l \Gamma(l)} e^{-\frac{x}{\theta}} . x^{l-1}$ $x \geq 0$ $\theta, l > 0$

(we have shown that $E(x) = l\theta$, $v(x) = I\theta^2$) .

Case (1) l is known , hence the p . d .f is

$F(x, \theta) = \frac{1}{\theta^l \Gamma(l)} e^{-\frac{x}{\theta}} . x^{l-1}$

$\log f(x, \theta) = -I \log \theta - \log(\Gamma(l)) - \frac{x}{\theta} + (l-1) \log x$

$\frac{d \log f\,(x,\theta)}{d\,\theta} = -\frac{l}{\theta} + \frac{x}{\theta^2}$

$E\,(\frac{d}{d\theta} \log f(x,\theta)) = -\frac{l}{\theta} + \frac{E\,(x)}{\theta^2} = 0 \rightarrow E(x) = l\theta$

$\frac{d2 \log f\,(x,\theta)}{d\,\theta^2} = \frac{l}{\theta^2} - \frac{2x}{\theta^3}$

$\therefore\ I\,(\theta) = E\,(-\frac{d\,2}{d\,\theta^{\,2}} \log(x,\theta)) = -\frac{l}{\theta^2} + \frac{2\,E\,(X)}{\theta^3} = \frac{l}{\theta^2}$

Alternatively:

$I\,(\theta) = E(\frac{dlog}{d\theta} f\,(x,\theta))^2 = E\,\{\frac{l^2}{\theta^2} + \frac{x^2}{\theta^4} - \frac{2\,x\,l}{\theta^3}\}$

$= \frac{l^2}{\theta^2} + \frac{l\,\theta^2 + l^2\theta^{\,2}}{\theta^4} - \frac{2l\,.l\,\theta}{\theta^3}$

$= \frac{l^2}{\theta^2} + \frac{l + l^2}{\theta^2} - \frac{2\,l^2}{\theta^2} = \frac{l}{\theta^2}$

Or rearrange we obtain: $E\,(\frac{x - l\,\theta}{\theta^2})^2 = \frac{l}{\theta^2}$ which is the same resultsas before .

Case (2) θ is known : the p . d .f .is $f(x, l) = \frac{1}{\theta^l\,\Gamma(l)} e^{-\frac{x}{\theta}} . x^{l-x}$

Log f (x , l) = $-l\,log\theta - \log \Gamma(l) - \frac{x}{\theta} + (l-1)logx$

$\frac{d \log f\,(x,l)}{d\,l} = -log\theta - \frac{d}{dl} \log\,(\Gamma(l)) + \log(x)$

= - log θ - F (l) – log x where F(l) = digammafunction

$\frac{d\,2 \log f\,(x,l)}{d\,l^{\,2}} = -\frac{d\,2}{d\,l^{\,2}} \log \Gamma\,(l) = -$ Ƒ(l) = Trigamma function

I (l) = E ($-\frac{d2}{d\,l^2} \log f(x, l)$) = Ƒ (l)

The digammafunction andTrigamma function are tabulated for several values of l , we have for large values of l as in " Whitker , Watson book , modern Analysis

F (E) $= \frac{d}{d\theta} \log\,\Gamma\,(l) = \log l - \frac{1}{2\,l}$

If l is known to be large , the information function can be approximated :

I (1) = F (1) = $\frac{d2}{d\,\,l^2} \log\,\,\Gamma(l) = \frac{d}{d\,l}$ F (1) = $\frac{1}{l} + \frac{1}{2l^2}$

Consider

I (1) = E ($\frac{d}{dl} \log f\,(x,l))^2$ = E { - log θ - F(l) + log x }^{2}

= E{(log θ)^2 +{F(e)}^{2} + (log x)^2 – 2 (log θ)(log x) – 2 F (l) log x + 2 (logθ) F (1) }.

= (log θ)^2 + { F(l) } ^{2} + E (log x) ^{2} – 2 log θ E (log x) – 2 F (l) E (log x) + 2 (log θ) F (l) .

We needthe following lemma :
Let the r. v .x have the gamma $dist^n$ with p .d .f

$$F(x) = \frac{1}{\theta^l\,\Gamma(l)} e^{-\frac{x}{\theta}} . x^{l-1}$$

Then

E (log x) = log θ + F (l)

E (log x) ^{2} = $\frac{1}{\Gamma(l)\theta^l}\frac{d}{dl}$ [Γ(l) θ^1] [log θ + F (l)]

proof : since f (x) is a p. d .f over 0 ≤ X≤ ∞ we must have $\int_{-\infty}^{\infty} f(x)dx = 1$

i .e $\int_0^{\infty} e^{-\frac{x}{\theta}} . x^{l-1} . dx = \theta^l \Gamma(l)$ …………………(A)

Differentiate equation (A) Both sides w. r. t. l , then

$\frac{d}{d\,l}\int_0^{\infty} e^{-\frac{x}{\theta}} x^{l-1}$ dx = $\int_0^{\infty} e^{-\frac{x}{\theta}}$ ($\frac{d}{dl} x^{l-1}$) dx

= $\int_0^{\infty} e^{-\frac{x}{\theta}} x^{l-1}$ log x dx

= $\frac{d}{dl}[\theta^l \Gamma(l)$

But $\frac{d}{dl}[\theta^l\Gamma(l) = \Gamma(l)\theta^l log\theta + \theta^l \frac{d}{d\,l}\Gamma(l)$

$= \Gamma(l)\theta^l\{log\theta + \frac{1}{\Gamma(l)}\frac{d}{dl}\,\Gamma(l)\}$

$= \Gamma(l)\theta^l\{log\theta + f(l)\}$

Hence $\int_0^{\theta} e^{-\frac{x}{\theta}}$ x^{l-1}(log x) dx = $\Gamma(l)\theta^l\{log\,\theta + F(l)\}$………………..(B)

Equation (B) can be written as:

$$\frac{1}{\Gamma(l)\theta^l}\int_0^{\infty} e^{-\frac{x}{\theta}} x^{l-1} \text{ (log x) dx} = \log\theta + F(l)$$

This mean E (logx) = log θ + F (l)

Similarly differentiate (B) w. r. t l we get :

$\int_0^{\infty} e^{-\frac{x}{\theta}} x^{l-1}$ (log x) ^{2} dx = $\frac{d}{dl}$ [$\Gamma(l)\theta^l$] [log θ + F (l)]

∴ E (log x) ^{2} = $\frac{1}{\Gamma(l)\theta^l}\frac{d}{dl}$ [Γ(l) θ^1] [log θ + F (l)]

Note : that we can show that

I (l) = E ($\frac{d}{dl}$ log $f(x, l)$) ^{2} = Ғ (l) Trigamma.

Fisher's information matrix :

Let the r.v .x have a p. d.f $f(x, \theta_1, \theta_2, \ldots\ldots, \theta_k)$

Depending on kunknown parameters $\theta_1, \theta_2, \ldots\ldots, \theta_k$. Again assume that we have a regular estimation case .

Since $\int_a^b f(x, \theta_1, \theta_2, \ldots\ldots\ldots, \theta_k)\, dx = 1$ …….. (1)

We get on differentiating both sides w. r. t . θi

$$\frac{d}{d\theta i}\int_a^b f\,(\,x\,,\theta_1,\theta_2,\dots,\theta_k\,)\,dx\; = 0 \;\to \int_a^b \frac{\mathrm{d}}{\mathrm{d}\theta\,\mathrm{i}}\mathrm{f}\,(\,x\,,\theta_1,\theta_2,\dots,\theta_k)dx = 0$$

i.e

$$\int_a^b \frac{\mathrm{d}}{\mathrm{d}\theta\,\mathrm{i}}\log \mathrm{f}(\;x\,,\theta_1,\theta_2,\dots,\theta_k)\;f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k)dx = 0 \quad \dots\dots\dots\dots(2)$$

This mean s

$$\mathrm{E}\,(\frac{d}{d\theta\;i}\log f\;(x\,,\theta_1,\theta_2,\dots,\theta_k)\;) = 0 \qquad i = 1,2,\dots,k \qquad \dots\dots\dots..(3)$$

Differentiating (2) w. r. t.θ_j j = 1 ,2, 3 ,…k we get :

$$\int_a^b (\frac{d2}{d\theta i\;d\theta j}\log f\;(x\,,\theta_1,\theta_2,\dots,\theta_k))\;f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k)\;dx + \int_a^b (\frac{d}{d\theta i}\log f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k)(\frac{d}{d\theta i}f\;(x\,,\theta_1,\theta_2,\dots,\theta_k) = \;0$$

Or

$$\int_a^b (\frac{d2}{d\theta i\;d\theta j}\log f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k))\;f\;(x\,,\theta_1,\theta_2,\dots,\theta_k)\;dx + \int_a^b (\frac{d}{d\theta i}\log f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k))(\frac{d}{d\theta i}f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k)f(x\,,\theta_1,\theta_2,\dots,\theta_k))dx = \;0$$

Hence

$$E(\frac{d}{d\theta i}\log\;f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k).\frac{d}{d\theta i}\log f\;(\,x\,,\theta_1,\theta_2,\dots,\theta_k)) = E\;(-\;\frac{d2}{d\theta_i\;d\theta_j}\log\;f(\,x\,,\theta_1,\theta_2,\dots,\theta_k))$$

$$\dots = Iij\;(\theta_1,\theta_2,\dots..,\theta_k)\dots\dots\dots(4)$$

Where : i=1,2,…….,k , j= 1,2,….,k.

Define a matrixI $(\theta_1,\theta_2,\dots,\theta_k) = ((I_{ij}))_{k*k}$

This is known a fisher's information matrix .

Remark (1) from eq (3) we have E $(\frac{d}{d\theta i}\log f(\,x\,,\theta_1,\theta_2,\dots,\theta_k) = 0\;then$

$$Iij = cov\{\frac{d}{d\theta i}\log f(\,x\,,\theta_1,\theta_2,\dots,\theta_k),\frac{d}{d\theta j}\log f(x\,,\theta_1,\theta_2,\dots,\theta_k)\}$$

$$= E\;\{\frac{d}{d\theta i}\log f(\,x\,,\theta_1,\theta_2,\dots,\theta_k),\frac{d}{d\theta i}\log f(x\,,\theta_1,\theta_2,\dots,\theta_k)\}$$

Remark (2) if $x_1, x_2, \dots, x_n$ is a random sample ofn independent observations on the x and if

$L = \prod_{i=1}^{n} f(x, \theta_1, \theta_2, \dots, \theta_k)$

Is thelikelihood function of the given sample , then we can show that

$E(\frac{d}{d\theta i} \log f(x_l, \theta_1, \dots\dots, \theta_k)) = 0$ I = 1,2, …., k

And

$E(-\frac{d2}{d\theta i\, d\theta j} \log l) = E\{\frac{d}{d\theta i} \log l \cdot \frac{d}{d\theta j} \log l\} = n\, I_{ij}$

Example(1) : consider the two – parameter normal $dist^n$.

$f(x, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$ $-\infty \le X \le \infty$, $\sigma^2 \ge 0$, $-\infty \le \theta \le \infty$

$\text{Log } f = -\log a - \log\sqrt{2\pi} - \frac{(x-\theta)^2}{2\sigma^2}$

$= -\log\sigma - const - \frac{x^2}{2\sigma^2} + \frac{\theta x}{\sigma^2} - \frac{\theta^2}{2\sigma^2}$

$$\frac{d}{d\theta}\log f = \frac{x}{\sigma^2} - \frac{\theta}{\sigma^2}$$

$E(\frac{d}{d\theta}\log f) = E(\frac{x}{\sigma^2} - \frac{\theta}{\sigma^2}) = 0$

$$\frac{d2 \log f}{d\,\theta^2} = -\frac{1}{\sigma^2}$$

$$\therefore Iij = E(-\frac{d2}{d\,\theta^2}\log f) = \frac{1}{\sigma^2}$$

Also$\frac{d2}{d\theta d\sigma^2}\log f = \frac{d2}{d\theta d\sigma}\log f\ \frac{da}{d\sigma^2} = \frac{1}{2\sigma}\frac{d2}{d\theta\ d\sigma}\log f$

$= \frac{1}{2\sigma}[-\frac{2x}{\sigma^3} + \frac{2\theta}{\sigma^3}] = \frac{-x+\theta}{\sigma^4}$

$\therefore I_{12} = E(-\frac{d2}{d\theta d\sigma^2}\log f) = 0$

$\frac{d}{d\sigma^2}\log f = \frac{1}{2\sigma}\frac{d}{d\sigma}\log f = \frac{1}{2\sigma}[-\frac{1}{\sigma} + \frac{x^2}{\sigma^3} - \frac{2\theta x}{\sigma^3} + \frac{\theta^2}{\sigma^3}]$

$= -\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4} - \frac{\theta x}{\sigma^4} + \frac{\theta^2}{2\sigma^4}$

$\therefore E(\frac{d}{d\sigma^2}\log f) = -\frac{1}{2\sigma^2} + \frac{E(x^2)}{2\sigma^4} - \frac{\theta\, E(X)}{\sigma^4} + \frac{\theta^2}{2\sigma^4}$

$= -\frac{1}{2\sigma^2} + \frac{E(x^2)}{2\sigma^4} - \frac{\theta\, E(X)}{\sigma^4} + \frac{\theta^2}{2\sigma^4} = 0$

$\frac{d2}{d^2\sigma^2}\log f = \frac{1}{2\sigma}\frac{d}{d\sigma}(\frac{d}{d\sigma^2}\log f)$

$= \frac{1}{2\sigma}\frac{d}{d\sigma}[-\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4} - \frac{\theta x}{\sigma^4} + \frac{\theta^2}{2\sigma^4}]$

$= \frac{1}{2\sigma}[\frac{1}{\sigma^3} + \frac{2x^2}{\sigma^5} - \frac{4\theta x}{\sigma^5} + \frac{2\theta^2}{\sigma^5}]$

$= \frac{1}{2\sigma^4} - \frac{x^2}{\sigma^6} + \frac{2\theta x}{\sigma^6} + \frac{\theta^2}{\sigma^6}$

$\therefore I_{12} = E(-\frac{d2}{d^2\sigma^2}\log f) = -\frac{1}{2\sigma^4} + \frac{\theta^2}{\sigma^6} + \frac{\sigma^2+\theta^2}{\sigma^6} - \frac{2\sigma^2}{\sigma^6}$

$= \frac{1}{2\sigma^4}$

Hence the information matrix is $I = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$

Example (2) : consider the two parameters gamma dist'n

$f(x, \theta, l) = \frac{1}{\theta^l \, \Gamma(l)} e^{-\frac{x}{\theta}} . x^{l-1}$

We have

$\text{Log } f = - l \log \theta - \log \Gamma(l) - \frac{x}{\theta} + (l - 1) \log x$

$\frac{d}{d\theta} \log f = - \frac{l}{\theta} + \frac{x}{\theta^2}$

$\frac{d2}{d\theta^2} \log f = \frac{l}{\theta^2} + \frac{2x}{\theta^3}$

$\therefore I_{11} = E\left(-\frac{d2}{d\theta^2} \log f \right) = \frac{l}{\theta^3}$

$\frac{d2}{d\theta \, dl} \log f = -\frac{1}{\theta}$

$\therefore I_{12} = E\left(-\frac{d2}{d\theta \, dl} \log f \right) = \frac{1}{\theta}$

$\frac{d}{dl} \log f = -log\theta - \mathcal{F}(l) + logx$

$\frac{d2}{dl^2} = -F(l)$

$\therefore I_{22} = E\left(-\frac{d2}{dl^2} \log f \right) = F(l)$

Hence the information matrix is

$I(\theta, l) = \begin{pmatrix} \frac{l}{\theta^2} & \frac{1}{\theta} \\ \frac{1}{\theta} & F(l) \end{pmatrix}$

References

- Frieden, B. Roy (2004) *Science from Fisher Information: A Unification.* Cambridge Univ. Press. ISBN 0-521-00911-1.
- Hald, A. (May 1999). "On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares". *Statistical Science* **14** (2): 214–222. JSTOR 2676741.
- Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley. ISBN 0-471-17912-4.
- Lehmann, E. L.; Casella, G. (1998). *Theory of Point Estimation* (2nd ed.). Springer. ISBN 0-387-98502-6.
- Pratt, John W. (May 1976). "F. Y. Edgeworth and R. A. Fisher on the Efficiency of Maximum Likelihood Estimation". *The Annals of Statistics* **4** (3): 501–514. doi:10.1214/aos/1176343457. JSTOR 2958222.
- Leonard J. Savage (May 1976). "On Rereading R. A. Fisher". *The Annals of Statistics* **4** (3): 441–500. doi:10.1214/aos/1176343456. JSTOR 2958221.
- Shrewish, Mark J. (1995). "Section 2.3.1". *Theory of Statistics*. New York: Springer. ISBN 0-387-94546-6.
- Stephen Stigler (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press. ISBN 0-674-40340-1.[*page needed*]
- Stephen M. Stigler (1978). "Francis YsidroEdgeworth, Statistician". *Journal of the Royal Statistical Society, Series A* **141** (3): 287–322. doi:10.2307/2344804. JSTOR 2344804.
- Stephen Stigler (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press. ISBN 0-674-83601-4.[*page needed*]
- Kendall, Maurice. And Stuart, Alan (1961). The Advance Theory Of Statistics. (Volume 2) .Charles Griffin & Company limited.
- Cramer, Harald. (1961). Mathematical Methods Of Statistics. Princeton University Press.
- Rao, Radhakrishna C.(1965). Linear Statistical Inference And Its Applications. (2nded) . John Wiley & Sons.