Efficiency for Cramer Rao and van Trees inequalities.

Gérard Letac, Université Paul Sabatier, Toulouse, France.

Conference at York, 01/30/08.

Fisher models Given a measurable space $(\Omega, \mathcal{A})$ and given an open subset Θ of $\mathbb{R}^k$ a Fisher model $(P_\theta(dw))_{\theta\in\Theta}$ is a model such that for all θ and $\theta' \in \Theta$, we have that P_θ is absolutely continuous with respect to $P_{\theta'}$. In other terms, P_θ and $P_{\theta'}$ are equivalent for all θ and $\theta' \in \Theta$.

Proposition 1. $(P_\theta(dw))_{\theta\in\Theta}$ is a Fisher model if and only if there exists a measure $\nu(dw)$ on $(\Omega, \mathcal{A})$ and a real function $(w, \theta) \mapsto \ell_w(\theta)$ such that

$$P_\theta(dw) = e^{\ell_w(\theta)}\nu(dw). \qquad (1)$$

Of course the pair $(\nu(dw), \ell_w(\theta))$ is not completely arbitrary since it satisfies

$$\int_\Omega P_\theta(dw) = 1 = \int_\Omega e^{\ell_w(\theta)}\nu(dw) \qquad (2)$$

Suppose now that $\theta \mapsto \ell_w(\theta)$ is differentiable and that there exists a positive and ν integrable function f such that $\|\ell'_w(\theta)|e^{\ell_w(\theta)} \leq f(w)$ for all $\theta \in \Theta$. In these circumstances we can differentiate under the sign integral and we get the important vector equality

$$\int_\Omega \ell'_w(\theta) P_\theta(dw) = 0 \qquad (3)$$

Some authors call $(w, \theta) \mapsto \ell'_w(\theta)$ the score function.

Examples.

The general exponential family, with $\ell_w(\theta) = \langle\theta, t(w)\rangle - k_\mu(\theta)$ where k_μ is the cumulant transform of the image $\mu(dx)$ of $\nu(dw)$ in $\mathbb{R}^k$ by $w \mapsto x = t(w)$.

Recall that a natural exponential family governed by μ where μ is a positive measure on $\mathbb{R}^k$ is the model on $\mathbb{R}^k$

$$P(\theta, \mu)(dx) = e^{\langle\theta,x\rangle - k_\mu(\theta)}\mu(dx)$$

where $\theta \in \Theta(\mu) \subset \mathbb{R}^k$, the l open convex set of existence of the Laplace transform of μ.

Recall that a general exponential family governed by (t, ν), where $t : \Omega \to \mathbb{R}^k$ and where ν is a positive measure on Ω is the model

$$P(\theta, t, \nu)(dw) = e^{\langle\theta,t(w)\rangle - k_\mu(\theta)}\nu(dw).$$

The location parameter model : If f is a strictly positive density in $\mathbb{R}^k$ take $\Omega = \Theta = \mathbb{R}^k$, $\nu(dw) = dw$ and $\ell_w(\theta) = \log f(x - \theta)$.

Fisher information Given a regular Fisher model $(P_\theta(dw))_{\theta\in\Theta}$ on Ω with the representation (1), the Fisher information is the (k,k) symmetric matrix

$$I(\theta)=\int_\Omega \ell'_w(\theta)\otimes\ell'_w(\theta)P_\theta(dw).$$

If $\theta\mapsto\ell_w(\theta)$ is twice differentiable and if conditions of differentiability under the integral are met, thus differentiating (3) once again gives

$$I(\theta)=-\int_\Omega \ell''_w(\theta)P_\theta(dw). \tag{4}$$

Note that $I(\theta)$ is obviously semi positive definite. We consider only Fisher models such that $(I(\theta)^{-1}$ does exist.

Under $P_\theta(dw)$ one can see $w\mapsto\ell'_w(\theta)$ as a random variable valued in $\mathbb{R}^k$. It is centered (by (3)) and $I(\theta)$ is its covariance matrix (by the definition of $I(\theta)$).

Examples.

The general exponential family : Here $I(\theta) = k''_\mu(\theta)$ by applying (4) and the fact that if $\ell_w(\theta) = \langle \theta, t(w) \rangle - k(\theta)$ then $\ell''_w = -k''_\mu$ which does not depend on $w \in \Omega$ and makes the computation of the integral trivial.

The location parameter model : denote $g = \log f$. Here the information matrix is the constant matrix

$$I(\theta) = \int_{\mathbb{R}^k} [g'(x) \otimes g'(x)] e^{g(x)} dx = -\int_{\mathbb{R}^k} g''(x) e^{g(x)} dx.$$

Examples : in the Gaussian case that is if $g(x) = -x^* \Sigma^{-1} x/2 - \frac{k}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma$ we get $I(\theta) = \Sigma^{-1}$. Sometimes g'' does not exist and only the definition of $I(\theta)$ can be used : example on the real line with the bilateral exponential distribution associated to $g(x) = -|x| - \log 2$. In this case $g'(x) = -\text{sign}\ x$, thus $g'^2(x) = 1$ and $I(\theta) = 1$.

Properties of the information matrix : changing parameter

Proposition 2. Let a regular Fisher model

$$(P_\theta(dw))_{\theta\in\Theta}$$

on Ω and consider a reparameterization

$$\Theta_1 \to \Theta, \quad t \mapsto f(t) = \theta$$

where f is a differentiable homeomorphism between the two open subsets Θ_1 and Θ of $\mathbb{R}^k$. Consider the Fisher model $(Q_t(dw))_{t\in\Theta_1}$ on Ω where $Q_t = P_{f(t)}$ and denote by $I_1(t)$ its information matrix. We have

$$I_1(t) = f'(t)^* I(f(t)) f'(t)$$

where $f'(t)$ is the Jacobian matrix of f and $f'(t)^*$ is its transpose matrix.

Example : The geometric distribution. For $\Omega = \mathbb{N}$ the set of non negative integers, consider the geometric distribution with three different parameterizations, thus three Fisher models with the same $\nu = \sum_{n\in\mathbb{N}} \delta_n$ and $k = 1$.

1. Classical parameterization : we take $\Theta_1 = (0, 1)$ and $P_p^{(1)}(dw) = (1-p)\sum_{n\in\mathbb{N}} p^n \delta_n(dw)$. Here $\ell_w(p) = w \log p + \log(1-p)$, $-\ell''_w(p) = \frac{w}{p^2} + \frac{1}{(1-p)^2}$ and the information is

$$I_1(p) = \frac{1}{p(1-p)^2}.$$

2. Canonical parameterization (in the sense of exponential families) : we take $\Theta_2 = (-\infty, 0)$ and $P_\theta^{(2)}(dw) = (1-e^\theta)\sum_{n\in\mathbb{N}} e^{\theta n}\delta_n(dw)$ leads to

$$I_2(\theta) = \frac{e^\theta}{(1-e^\theta)^2}.$$

3. Parameterization by the mean : we take $\Theta_3 = (0, \infty)$ and $P_m^{(3)}(dw) = \sum_{n\in\mathbb{N}} \frac{m^n}{(1+m)^{n+1}}\delta_n(dw)$ leads to

$$I_3(m) = \frac{1}{m + m^2}.$$

Example : The general exponential family. We reparameterize it by the mean $m = k'_\mu(\theta)$. Here $\Theta_1 = M_F$ is the domain of the means of the natural exponential family $F = F(\mu)$ which is associated to the general exponential family, and f is the inverse function $m \mapsto \theta = \psi_\mu(m)$ of $\theta \mapsto m = k'_\mu(\theta)$. Thus $\psi'_\mu(m)$ is $V_F(m)^{-1}$ which is the variance function of the NEF F. Recall that $I(\theta) = k''_\mu(\theta)$ and thus $I(\psi_\mu(m)) = V_F(m)$. Finally Proposition 2 implies that $I_1(m) = V_F(m)^{-1}$ a strong contrast with $I(\psi_\mu(m)) = V_F(m)$. Note that $\ell'_w(\theta) = t(w) - k'_\mu(\theta)$ and therefore if $h_w(m) = \ell_w(\psi_\mu(m))$ we get

$$h'_w(m) = V_F(m)^{-1}(t(w) - m). \qquad (5)$$

Properties of the information matrix : independence

Proposition 3. Consider two regular Fisher models $(P_\theta(dw))_{\theta\in\Theta}$ on Ω and $(Q_\theta(dw_1))_{\theta\in\Theta}$ on Ω_1, with information matrices $I(\theta)$ and $I_1(\theta)$. Then the information matrix of the product Fisher model $(P_\theta\otimes Q_\theta)(dw, dw_1))_{\theta\in\Theta}$ is $I(\theta)+I_1(\theta)$.

Proof. Use the representations $P_\theta(dw) = e^{\ell_w(\theta)}\nu(dw)$ and $Q_\theta(dw_1) = e^{h_{w_1}(\theta)}\mu(dw_1)$. This leads to the representation

$$(P_\theta\otimes Q_\theta)(dw, dw_1) = e^{\ell_w(\theta)+h_{w_1}(\theta)}(\nu\otimes\mu)(dw, dw_1).$$

If we assume the existence of second derivatives, things are simple since the information of the product is

$$\int_\Omega\int_{\Omega_1}[\ell''_w(\theta)+h''_{w_1}(\theta)]P_\theta(dw)Q_\theta(dw_1) = I(\theta)+I_1(\theta)$$

As a consequence, suppose that $X_1, \ldots, X_N$ are iid random variables valued in Ω with the same distribution $P_\theta(dw)$, being a Fisher model with information matrix $I(\theta)$. Then the information matrix for the Fisher model on Ω^N which is the distribution of $(X_1, \ldots, X_N)$ is $NI(\theta)$.

The Cramér-Rao inequality. The most elegant version is this

Proposition 4. Consider a regular Fisher model $(P_\theta(dw))_{\theta\in\Theta}$ on Ω. Let $X(w)$ be an unbiased estimator X of θ. Then the matrix

$$\begin{bmatrix} \mathrm{Cov}_\theta(X) & I_k \\ I_k & I(\theta) \end{bmatrix} \tag{6}$$

is semipositive definite. In particular when $(I(\theta))^{-1}$ exists, this is equivalent to saying that

$$\mathrm{Cov}_\theta(X) - (I(\theta))^{-1}$$

is semipositive definite and this is called the *Cramér-Rao inequality.* In particular,

$$\mathrm{Cov}_{\theta_0}(X) - (I(\theta_0))^{-1} = 0$$

if and only if ν almost everywhere we have $X(w) = \theta_0 + I(\theta_0)^{-1}\ell'_w(\theta_0)$ with the representation (1) of the Fisher model. Finally, if θ is randomized and has distribution $\lambda(d\theta)$ we have the *Bayesian Cramér-Rao inequality* which is the following statement :

$$\int_\Theta \mathrm{Cov}_\theta(X)\lambda(d\theta) - (\int_\Theta I(\theta)\lambda(d\theta))^{-1} \tag{7}$$

is semi positive definite.

Proof. The hypothesis implies

$$\theta = \int_\Omega X(w) P_\theta(dw) = \int_\Omega X(w) e^{\ell_w(\theta)} \nu(dw).$$

Taking derivative of both sides gives

$$I_k = \int_\Omega X(w) \otimes \ell'_w(\theta) e^{\ell_w(\theta)} \nu(dw) = \int_\Omega X(w) \otimes \ell'_w(\theta) P_\theta(dw)$$

Recall that $0 = \int_\Omega \ell'_w(\theta) P_\theta(dw)$ and thus $0 = \int_\Omega \theta \otimes \ell'_w(\theta) P_\theta(dw)$. We get from this :

$$I_k = \int_\Omega (X(w) - \theta) \otimes \ell'_w(\theta) P_\theta(dw) \qquad (8)$$

Since $(X(w) - \theta)$ and $\ell'_w(\theta)$ are centered random variables for the probability $P_\theta(dw)$ the equality (8) says that the covariance of these two random variables is I_k. Finally the covariance matrix of the random variable

$$(X(w) - \theta, \ell'_w(\theta))$$

of $\mathbb{R}^{2k}$ is given by (6) and thus (6) is semi-positive definite.

To finish the proof, we observe that

$$\begin{bmatrix} I_k & -(I(\theta))^{-1} \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \mathrm{Cov}_\theta(X) & I_k \\ I_k & I(\theta) \end{bmatrix} \begin{bmatrix} I_k & 0 \\ -(I(\theta))^{-1} & I_k \end{bmatrix}$$
$$= \begin{bmatrix} \mathrm{Cov}_\theta(X) - (I(\theta))^{-1} & 0 \\ 0 & I(\theta) \end{bmatrix}$$

and this shows that $\mathrm{Cov}_\theta(X) - (I(\theta))^{-1}$ is semipositive definite. Finally for seeing that

$$\mathrm{Cov}_{\theta_0}(X)-(I(\theta_0))^{-1} = 0 \Leftrightarrow X(w) = \theta_0 + I(\theta_0)^{-1}\ell'_w(\theta_0)$$

we compute the covariance matrix of

$$\begin{bmatrix} I_k & -(I(\theta))^{-1} \\ 0 & I_k \end{bmatrix} \begin{bmatrix} X(w) - \theta \\ \ell'_w(\theta) \end{bmatrix} =$$

$$\begin{bmatrix} X(w) - \theta - (I(\theta))^{-1}\ell'_w(\theta) \\ \ell'_w(\theta) \end{bmatrix}$$

and equality (9) shows that this covariance is

$$\begin{bmatrix} \mathrm{Cov}_\theta(X) - (I(\theta))^{-1} & 0 \\ 0 & I(\theta) \end{bmatrix}$$

If $\mathrm{Cov}_{\theta_0}(X) - (I(\theta_0))^{-1} = 0$ the covariance is $\begin{bmatrix} 0 & 0 \\ 0 & I(\theta_0) \end{bmatrix}$. Thus $X(w) = \theta_0 + I(\theta_0)^{-1}\ell'_w(\theta_0)$ this equality being P_{θ_0} everywhere, thus $\nu(dw)$ everywhere.

For the last part, we use the fact that sums or convex combinations of semi positive definite matrices are also semipositive definite, and thus

$$\begin{bmatrix} \int_\Theta \mathrm{Cov}_\theta(X)\lambda(d\theta) & I_k \\ I_k & \int_\Theta I(\theta)\lambda(d\theta) \end{bmatrix} \tag{10}$$

is semi positive definite which implies (7).

Efficiency. Proposition 4 leads to a characterization of the general exponential families, which are the only Fisher models able to realize always equality in the above Cramér-Rao inequality.

Proposition 5. Consider a regular Fisher model $(P_\theta(dw))_{\theta\in\Theta}$ on Ω and an unbiased estimator X of θ. Assume that the image $\mu(dx)$ of $\nu(dw)$ by the map $w \mapsto x = X(w)$ is not concentrated on an affine hyperplane of $\mathbb{R}^k$. Then $\text{Cov}_\theta(X) - (I(\theta))^{-1} = 0$ for all $\theta \in \Theta$ if and only if $(P_\theta(dw))_{\theta\in\Theta}$ is a general exponential family parameterized by its mean. Under these circumstances $X(w) = t(w)$.

Proof.

$\Leftarrow$ If we parameterize a general exponential family by the mean, and if $X(w) = t(w)$ then the equality $\mathrm{Cov}_m(X) - (I(m))^{-1} = 0$ takes place. To see this apply the definitions, the fact that the information is $V_F(m)^{-1}$ and (7).

$\Rightarrow$ It will be easier to follow if we rather denote the parameter by m and the parameter set by M instead of θ and Θ. The Fisher model is $P_m(dw) = e^{h_w(m)}\nu(dw)$ and we denote the information by $I(m)$. From Proposition 4 the hypothesis $\mathrm{Cov}_m(X) - (I(m))^{-1} = 0$ for all $m \in M$ is translated into $X(w) = m + I(m)^{-1}h'_w(m)$ for each m and ν almost everywhere. Thus $h'_w(m) = I(m)(X(w) - m)$. Since this is true for all $X(w)$ this implies that both $I(m)$ and $I(m)m$ are gradients and thus there exists $A : M \to \mathbb{R}^k$ such that $A' = I$ and there exists $B : M \to \mathbb{R}$ such that $B'(m) = I(m)m$. Thus $h_w(m) = \langle A(m),\ X(w)\rangle - B(m)$.

Since $\int_\Omega P_m(dw) = 1$ we have

$$\int_\Omega e^{\langle A(m),\, X(w)\rangle} \nu(dw) = e^{B(m)}.$$

Since $A'(m) = I(m)$ is never singular the map $m \mapsto \theta = A(m)$ can be locally inverted in a neighborhood U_0 of any point m_0 and denote $A(U_0) = \Theta_0$. Thus for all $\theta \in \Theta_0$ we have

$$\int_{\mathbb{R}^k} e^{\langle \theta,\, x\rangle} \mu(dx) = e^{k(\theta)}.$$

and thus for $m \in U_0$ we have $B(m) = k(A(m))$. Taking differentials : $I(m)m = B'(m) = I(m)k'_\mu(A(m))$ which implies $k'_\mu(A(m)) = m$ and $k'_\mu(\theta) = m$ which shows that the natural exponential family generated by μ admits m as the mean parameter. This is true for any m_0 and the result is proved.

Example : Characterization of the Gaussian distribution. In the location parameter model $\{f(x-\theta)\,;\theta \in \mathbb{R}^k\}$ we assume that $\int_{\mathbb{R}^k} xf(x)dx = 0$ thus $X(x) = x$ is an unbiased estimator of θ. Now, assume that Cov$(X) - I^{-1} = 0$ (recall that Cov$_\theta(X)$ and $I(\theta)$ are constants in the location parameter model). Proposition 5 implies that the model is a general exponential family F such that $t(x) = x$, thus F is actually a natural exponential family (NEF). Furthermore, this NEF is invariant by translation, thus its variance function is a constant Σ, thus this NEF is the family of Gaussian distributions in $\mathbb{R}^k$ with known covariance Σ and unknown mean.

Finally, we give the general Cramér-Rao inequality, a little less elegant than Proposition 4, but which contains it.

Proposition 6. Consider a regular Fisher model $(P_\theta(dw))_{\theta\in\Theta}$ on Ω, with $\Theta \subset \mathbb{R}^k$. Let $w \mapsto X(w)$ be a measurable map from Ω to $\mathbb{R}^m$ (where m is not necessarily equal to k). Denote $\psi(\theta) = \int_\Omega X(w)P_\theta(dw)$ and assume that differentiability conditions in the integral are met. Introduce the differential $\psi'(\theta) : \mathbb{R}^k \to \mathbb{R}^m$ and its transpose $\psi'(\theta)^* : \mathbb{R}^m \to \mathbb{R}^k$. Then the matrix

$$\begin{bmatrix} \mathrm{Cov}_\theta(X) & \psi'(\theta) \\ \psi'(\theta)^* & I(\theta) \end{bmatrix} \tag{11}$$

is semipositive definite. In particular when $(I(\theta))^{-1}$ exists, this is equivalent to saying that

$$\mathrm{Cov}_\theta(X) - \psi'(\theta)(I(\theta))^{-1}\psi'(\theta)^*$$

is semipositive definite. In particular, $\mathrm{Cov}_{\theta_0}(X) - \psi'(\theta_0)(I(\theta_0))^{-1}\psi'(\theta_0)^* = 0$ if and only if ν almost everywhere we have $X(w) = \psi(\theta_0) + \psi'(\theta_0)I(\theta_0)^{-1}\ell'_w(\theta_0)$.

Proof. It suffices to see that $\mathrm{Cov}_\theta(X)-\psi'(\theta)(I(\theta))^{-1}\psi'(\theta$ is the covariance of $[X(w)-\psi(\theta), \ell'_w(\theta)]$, a random vector of $\mathbb{R}^{k+m}$. This is done in an entirely similar way as in Proposition 4. For studying the equality case we observe that

$$\begin{bmatrix} I_m & -\psi'(\theta)(I(\theta))^{-1} \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \mathrm{Cov}_\theta(X) & \psi'(\theta) \\ \psi'(\theta)^* & I(\theta) \end{bmatrix} \begin{bmatrix} I_m \\ -(I(\theta))^{-1}\psi' \end{bmatrix}$$

is equal to $\begin{bmatrix} \mathrm{Cov}_\theta(X)-\psi'(\theta)(I(\theta))^{-1}\psi'(\theta)^* & 0 \\ 0 & I(\theta) \end{bmatrix}$ which is the covariance of the random vector of $\mathbb{R}^{k+m}$ defined by

$$[X(w)-\psi(\theta)-\psi'(\theta)I(\theta)^{-1}\ell'_w(\theta), \ell'_w(\theta)].$$

Thus if $\mathrm{Cov}_{\theta_0}(X)-\psi'(\theta_0)(I(\theta_0))^{-1}\psi'(\theta_0)^*=0$ we have as claimed the following equality ν almost everywhere :

$$X(w)=\psi(\theta_0)+\psi'(\theta_0)I(\theta_0)^{-1}\ell'_w(\theta_0).$$

Examples. If $m = 1$, Proposition 6 is the classical Cramér-Rao inequality

$$\mathrm{Var}_\theta(X) \geq \psi'(\theta)(I(\theta))^{-1}\psi'(\theta)^*.$$

If furthermore $k = 1$ it can be read as

$$\mathrm{Var}_\theta(X) \geq \frac{\psi'(\theta)^2}{I(\theta)}.$$

This is the original form found by Fréchet (1938), Cramér (1945) page 475 and C. R. Rao (1946) and proved in elementary courses as follows :

$$\begin{aligned}
\psi(\theta) &= \int_\Omega X(w)P_\theta(dw), \\
\psi'(\theta) &= \int_\Omega X(w)\ell'_w(\theta)P_\theta(dw), \\
0 &= \int_\Omega \ell'_w(\theta)P_\theta(dw), \\
\psi'(\theta) &= \int_\Omega [X(w) - \psi(\theta)]\ell'_w(\theta)P_\theta(dw)
\end{aligned}$$

and we use Schwarz inequality :

$$[\psi'(\theta)]^2 \leq \int_\Omega [X(w)-\psi(\theta)]^2 P_\theta(dw) \int_\Omega [\ell'_w(\theta)]^2 P_\theta(dw)$$

$$= \mathrm{Var}_\theta(X)I(\theta)$$

Density information and the van Trees inequality. Suppose that $\lambda(d\theta)$ is a C_1 probability distribution on the open subset Θ of $\mathbb{R}^k$ which is > 0 on Θ and write $\lambda(\theta) = e^{g(\theta)}$. The symmetric matrix of order k

$$I_\lambda = \int_\Theta [g'(\theta) \otimes g'(\theta)] e^{g(\theta)} d\theta \qquad (12)$$

is called the density information of λ. This is not in general the Fisher information of some Fisher model, except when $\Theta = \mathbb{R}^k$ where I_λ is the Fisher information of the location parameter model defined by

$$P_\theta(dx) = \lambda(x - \theta)dx.$$

Examples of density information.

Non singular normal distributions. If $\lambda(\theta)d\theta$ is $N(m, \Sigma)$ in $\mathbb{R}^k$, the density information matrix is $I_\lambda = \Sigma^{-1}$.

Densities invariant by rotation. We assume that $g(\theta) = g_1(r)$ where $r = \|\theta\|$ with the canonical Euclidean structure on $\mathbb{R}^k$. In this case the density information is I_k multiplied by the scalar number

$$\frac{\pi^{k/2}}{\Gamma(1+\frac{k}{2})}\int_0^\infty g_1'(r)^2 e^{g_1(r)} r^{k-1} dr.$$

Consider the following example

$$\lambda(d\theta) = \frac{\Gamma(\frac{k}{2}+b)}{\pi^{k/2}\Gamma(b)}(1-\|\theta\|^2)_+^{b-1}d\theta$$

where x_+ means x if $x>0$ and 0 if $x \leq 0$. Here $g_1'(r) = -(b-1)\frac{2r}{1-r^2}$. To see that λ is indeed a probability on the unit ball of $\mathbb{R}^k$ observe that

$$\frac{2\pi^{k/2}}{\Gamma(\frac{k}{2})} \times \frac{\Gamma(\frac{k}{2}+b)}{\pi^{k/2}\Gamma(b)}\int_0^1 r^{k-1}(1-r^2)^{b-1}dr = 1.$$

Thus the scalar factor of I_k of the density information is

$$\frac{\pi^{k/2}}{\Gamma(1+\frac{k}{2})}\frac{\Gamma(\frac{k}{2}+b)}{\pi^{k/2}\Gamma(b)}\times 4(b-1)^2\int_0^1 r^{k+1}(1-r^2)^{b-3}dr.$$

This information exists only if $b>2$ and is

$$I_\lambda = \frac{2}{b}(b-1)(b-1+\frac{k}{2})I_k.$$

Proposition 7. Let $(P_\theta)_{\theta\in\Theta}$ be a regular Fisher model on Ω where Θ is an open subset of $\mathbb{R}^k$. Denote its Fisher information by $I(\theta)$. Let $\lambda(\theta)$ be a probability density on $\overline{\Theta}$ which is C_1 and > 0 on Θ and which is zero on the boundary $\partial\Theta = \overline{\Theta}\setminus\Theta$ and at infinity. Denote its density information by I_λ. Let $\Omega \ni w \mapsto X(w) \in \mathbb{R}^k$ be an arbitrary estimator of θ. Write

$$C = \int_{\Omega\times\Theta}[(X(w)-\theta)\otimes(X(w)-\theta)]P_\theta(dw)\lambda(\theta)d\theta$$

Under these circumstances the following $2k$ symmetric matrix

$$\begin{bmatrix} C & I_k \\ I_k & I_\lambda + \int_\Theta I(\theta)\lambda(\theta)d\theta \end{bmatrix} \tag{13}$$

is semi positive definite. In particular the k symmetric matrix

$$\int_{\Omega\times\Theta}[(X(w)-\theta)\otimes(X(w)-\theta)]P_\theta(dw)\lambda(\theta)d\theta$$

$$-[I_\lambda + \int_\Theta I(\theta)\lambda(\theta)d\theta]^{-1}$$

is semi positive definite (van Trees inequality) provided that the matrix $I_\lambda + \int_\Theta I(\theta)\lambda(\theta)d\theta$ is invertible.

Proof. Denote $P_\theta(dw) = e^{\ell_w(\theta)}\nu(dw)$ as usual and write $\lambda(\theta) = e^{g(\theta)}$. The basic trick of the proof is the Stokes' theorem, which says that if f is a function on $\overline{\Theta}$ valued in $\mathbb{R}^m$ which is sufficiently regular and which is zero on $\partial\Theta$ then

$$\int_\Theta f'(\theta)d\theta = 0.$$

We apply this principle to the two following functions

1. $f(\theta) = \lambda(\theta)e^{\ell_w(\theta)}$ and thus $m = 1$. Here
$$f'(\theta) = (g'(\theta) + \ell'_w(\theta))\lambda(\theta)e^{\ell_w(\theta)} \qquad (14)$$
2. $f(\theta) = \theta\lambda(\theta)e^{\ell_w(\theta)}$ and thus $m = k$. Here
$$f'(\theta) = [I_k + \theta \otimes (g'(\theta) + \ell'_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)} \qquad (15)$$

From (14) and Stokes we get that $X(w) \otimes \int_\Theta (g'(\theta) + \ell'_w(\theta))\lambda(\theta)e^{\ell_w(\theta)}d\theta = 0$ and thus integrating with respect to $\nu(dw)$ we have

$$\int_{\Omega\times\Theta}[(X(w)\otimes(g'(\theta)+\ell'_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)}\nu(dw)d\theta = 0. \tag{16}$$

From (15) and Stokes we get that

$$\int_\Theta[\theta\otimes(g'(\theta)+\ell'_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)}d\theta = -I_k\int_\Theta \lambda(\theta)e^{\ell_w(\theta)}d\theta$$

and thus integrating with respect to $\nu(dw)$ we have

$$\int_{\Omega\times\Theta}[\theta\otimes(g'(\theta)+\ell'_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)}\nu(dw)d\theta = -I_k. \tag{17}$$

since $P(d\theta, dw) = \lambda(\theta)e^{\ell_w(\theta)}\nu(dw)d\theta$ is a probability on $\Omega\times\Theta$. We now combine (16) and (17) to get finally

$$\int_{\Omega\times\Theta}[(X(w)-\theta)\otimes(g'(\theta)+\ell'_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)}\nu(dw)d\theta = I_k.$$

Now consider the expectation under $P(d\theta, dw)$ of the random variable $(\theta, w) \mapsto$

$$[(X(w)-\theta),(g'(\theta)+\ell'_w(\theta)]\otimes[(X(w)-\theta),(g'(\theta)+\ell'_w(\theta)]$$

which is valued in the set of semi positive definite matrices of order $2k$. This expectation is nothing but the matrix (13) which is therefore semi positive definite also. To prove this the only point left is to check that

$$\int_{\Omega\times\Theta}[(g'(\theta)+\ell_w(\theta))\otimes(g'(\theta)+\ell_w(\theta))]\lambda(\theta)e^{\ell_w(\theta)}\nu(dw)d\theta =$$

This comes from

$$\int_\Omega \ell'_w(\theta)e^{\ell_w(\theta)}\nu(dw) = 0 \Rightarrow \int_\Omega[g'(\theta)\otimes\ell'_w(\theta)]e^{\ell_w(\theta)}\nu(dw) =$$

Finally, passing from (13) to the proper van Trees inequality is standard and is done in the same way as done for the Cramér-Rao inequality.

Efficiency in the van Trees inequality

Proposition 8. We keep the hypothesis and the notations of Proposition 8. Denote

$$\Sigma = [I_\lambda + \int_\Theta I(\theta)\lambda(\theta)d\theta]^{-1}.$$

Then

$$\int_{\Omega\times\Theta}[(X(w)-\theta)\otimes(X(w)-\theta)]P_\theta(dw)\lambda(\theta)d\theta-\Sigma = 0$$

if and only if the Fisher model is a general exponential family on Ω generated by $t(w) = \Sigma^{-1}X(w)$ and by a measure $\nu_1(dw)$ equivalent to ν. Under this circumstance, if $\mu(dx)$ is the image in $\mathbb{R}^k$ of $\nu_1(dw)$ by $w \mapsto t(w)$ with Laplace transform $L_\mu = e^{k_\mu}$ the density λ is

$$\lambda(\theta) = L_\mu(\theta)e^{-\frac{1}{2}\langle\theta,\Sigma^{-1}(\theta)\rangle} = e^{-\frac{1}{2}\langle\theta,\Sigma^{-1}(\theta)\rangle+k_\mu(\theta)}.$$

Furthermore $\mu_1(dx) = e^{\frac{1}{2}\langle x,\,\Sigma^{-1}x\rangle}\mu(dx)$ is a probability and λ is the convolution $N(0,\Sigma)*\mu_1$.

Proof. The hypothesis is saying that $P(d\theta, dw)$ almost surely we have

$$X(w) - \theta = \Sigma[g'(\theta) + \ell'_w(\theta)].$$

This can be seen as in Proposition 5, after multiplication of the matrix of order $2k$ of (13) by a suitable triangular matrix. As a consequence we have

$$\Sigma^{-1}X(w) - \ell'_w(\theta) = g'(\theta) + \Sigma^{-1}\theta.$$

Integrating, there exists a function $C(w)$ such that

$$\langle\theta, \Sigma^{-1}X(w)\rangle - \ell_w(\theta) = g(\theta) + \frac{1}{2}\langle\theta, \Sigma^{-1}\theta\rangle + C(w).$$

and finally

$$\lambda(\theta)e^{\ell_w(\theta)} = e^{\langle\theta,\, \Sigma^{-1}X(w)\rangle - \frac{1}{2}\langle\theta,\, \Sigma^{-1}\theta)\rangle - C(w)}.$$

Now we multiply both sides by $\nu(dw)$, we denote $\nu_1(dw) = e^{-C(w)}\nu(dw)$ we integrate in w and we use the equality (2) to claim that

$$\begin{aligned}
\lambda(\theta) &= \int_\Omega e^{\langle\theta,\,\Sigma^{-1}X(w)\rangle-\frac{1}{2}\langle\theta,\,\Sigma^{-1}\theta)\rangle}\nu_1(dw)\\
&= e^{-\frac{1}{2}\langle\theta,\,\Sigma^{-1}\theta)\rangle}\int_{\mathbb{R}^k} e^{\langle\theta,x\rangle}\mu(dx)\\
&= L_\mu(\theta)e^{-\frac{1}{2}\langle\theta,\Sigma^{-1}(\theta)\rangle}\\
&= e^{-\frac{1}{2}\langle\theta,\Sigma^{-1}(\theta)\rangle+k_\mu(\theta)}\\
&= \int_{\mathbb{R}^k} e^{-\frac{1}{2}\langle\theta-x,\,\Sigma^{-1}(\theta-x))\rangle}e^{\frac{1}{2}\langle x,\,\Sigma^{-1}x\rangle}\mu(dx)
\end{aligned}$$

Why do we have $\Theta = \mathbb{R}^k$? If not there exists $\theta_0 \in \partial\Theta$. Thus on this point θ_0, in order to fill the condition $\lim_{\theta\to\theta_0}\lambda(\theta) = 0$ we have $\lim_{\theta\to\theta_0} -\frac{1}{2}\langle\theta,\Sigma^{-1}(\theta)\rangle + k_\mu(\theta) = -\infty$ and thus $\lim_{\theta\to\theta_0} k_\mu(\theta) = -\infty$ which is impossible since k_μ is convex.

As a consequence the Fisher model is

$$e^{\ell_w(\theta)}\nu(dw) = e^{\langle\theta,\,\Sigma^{-1}X(w)\rangle-k_\mu(\theta)}\nu_1(dw)$$

which is a general exponential family as claimed.

Restriction of the location parameter model to a subspace. This example is motivated by a correspondence with Abram Kagan who studies with his students the case of a two dimensional density with one dimensional unknown parameter θ of the form

$$\{f(x-\theta, y-\theta)\ ;\ \theta \in \mathbb{R}\}.$$

This is a Fisher model on $\Omega = \mathbb{R}^2$ if f is differentiable and positive. If furthermore

$$\int_{\mathbb{R}^2} xf(x,y)dxdy = \int_{\mathbb{R}^2} yf(x,y)dxdy = 0$$

clearly $aX + (1-a)Y$ is an unbiased estimator of θ for any $a \in \mathbb{R}$. We abstract this situation by replacing $\mathbb{R}^2$ by an arbitrary linear space E, by replacing the linear subspace $H = \{(\theta, \theta)\ ;\ \theta \in \mathbb{R}\}$ of $\mathbb{R}^2$ by an arbitrary subspace H of E, by replacing the projection

$$(x, y) \mapsto (ax + (1-a)y, ax + (1-a)y)$$

from $\mathbb{R}^2$ to H of kernel

$$H_1 = \{(x, y)\ ;\ ax + (1-a)y = 0\}$$

by an arbitrary projection on H.

Let E be a real linear space of dimension k. Let $\{f(x-\theta)dx\ ;\ \theta \in E\}$ be a location parameter model and let H be a linear subspace of E. We do not use any notion of orthogonality on E. Thus we denote by $H^{\perp}$ the subspace of the elements $\alpha \in E^*$ such that $\alpha(x) = 0$ for all $x \in H$. It is classical that H^* is canonically isomorphic to $E^*/H^{\perp}$.

Consider the Fisher model

$$\{f(x-\theta)dx\ ;\ \theta \in H\}.$$

Since we assume $f > 0$ we write $f = e^g$ as usual. We have $\ell_x(\theta) = g(x-\theta)$ and therefore $\ell'_x(\theta)$ is an element of H^*. Its realisation from the differential of g is subtle, since $g'(x)$ is an element of E^*. Actually $\ell'_x(\theta)$ is the image of $-g'(x-\theta)$ in H^* by the canonical projection π from E^* onto $E^*/H^{\perp} = H^*$.

Thus the information

$$
\begin{aligned}
I(\theta) &= \int_E [\ell'_x(\theta) \otimes \ell'_x(\theta)] f(x-\theta) dx \\
&= \int_E [\pi(g'(x)) \otimes \pi(g'(x))] f(x) dx
\end{aligned}
$$

is an element of $\mathcal{L}_s(H, H^*)$. It does not depend on θ.

We select a subspace H_1 of E which is complementary of H, *i.e.* such that $H + H_1 = E$ and $H \cap H_1 = \{0\}$. We assume $\int_E x f(x) dx = 0$. We denote by π_1 the projection from E to H which is parallel to H_1 and we consider the estimator $X_1 = \pi_1(X)$ of θ where $X \sim f(x-\theta)dx$. Of course X is unbiased by the hypothesis $\int_{\mathbb{R}^k} x f(x) dx = 0$ and by linearity X_1 is unbiased too. Now we prove the following

Proposition 9. Suppose that the unbiased estimator X_1 realizes equality in the Cramér-Rao inequality, that means $\mathrm{Cov}_\theta(X_1) - (I(\theta))^{-1} = 0$ for all $\theta \in H$. Under such a circumstance there exists $\Sigma \in \mathcal{L}_s(H^*, H)$ which is positive definite and a probability density $d(h_1)$ on H_1 such that

$$f(h, h_1) = Ce^{-\frac{1}{2}\langle \Sigma^{-1}(h),\, h\rangle} d(h_1).$$

Proof. It is convenient to write $E = H \times H_1$ and to denote the elements of E by $(h, h_1) = x$, with $\pi_1(x) = h$. We rather write m instead of θ. Thus the model is

$$\{f(h-m, h_1)dhdh_1 \ ; \ m \in H\}.$$

From Proposition 5 where (Ω, E, X, w) is replaced by (E, H, X_1, x) we can claim that the model is a general exponential family parameterized by the mean m. Thus it has the form

$$f(h-m, h_1)dhdh_1 = e^{\langle\psi_\mu(m),h\rangle - k_\mu(\psi_\mu(m))}\nu(dh, dh_1)$$

for some unknown measure ν on E and where μ is the image of ν by the projection $(h, h_1) \mapsto h$. The above equality shows that ν is absolutely continuous with respect to dx and that the converse is true : thus $\nu(dh, dh_1) = e^{t(h,h_1)}dhdh_1$. We prefer to write $t(h, h_1) = t_{h_1}(h)$ since we are going to use derivative with respect to h only, and this avoids ∂ symbols. Writing also $f(h, h_1) = e^{g_{h_1}(h)}$ we finally get

$$g_{h_1}(h-m) = \langle\psi_\mu(m), h\rangle - k_\mu(\psi_\mu(m)) + t_{h_1}(h).$$

Taking differential of this expression with respect to h we get

$$g'_{h_1}(h-m) = \psi_\mu(m) + t'_{h_1}(h).$$

This shows that $h \mapsto g'_{h_1}(h)$ is a linear function from H to $\mathbb{R}$ thus $g'_{h_1}(h) \in H^*$. In principle this linear function depends on h_1. However doing $h = 0$ shows that neither g'_{h_1} nor t'_{h_1} depend on h_1. Finally $m \mapsto \psi_\mu(m)$ is linear. Let us write it $\psi_\mu(m) = \Sigma^{-1}(m)$ where $\Sigma \in \mathcal{L}_s(H^*, H)$. Thus $t'_{h_1}(h) = -\Sigma^{-1}(h) = g'_{h_1}(h)$. Integrating there exists a function $r(h_1)$ such that $g_{h_1}(h) = -\frac{1}{2}\langle \Sigma^{-1}(h), h\rangle + r(t_1)$. This proves the desired result.