What can be discovered in a formula for the variance of the BLUE

Andrej Pázman

Department of Applied Mathematics and Statistics
Comenius University Bratislava

Winter Workshop on Mathematical Statistics 2008
December 15-16
Bratislava

Results from textbooks

Linear model:

$$\begin{aligned} y &= F\theta + \varepsilon; \qquad \theta \in \mathbb{R}^p \\ E(\varepsilon) &= 0, Var(\varepsilon) = \sigma^2 I \end{aligned}$$

the variance of the BLUE for $c^T\theta$ is

$$Var\left(c^T\hat{\theta}\right) = \sigma^2 c^T M^- c = \sigma^2 \max_{\alpha: M\alpha \neq 0} \frac{\left(c^T\alpha\right)^2}{\alpha^T M \alpha}$$

where $M = F^T F$ is the information matrix
(cf. textbooks of J. Anděl or of C.R. Rao)

Results from textbooks

Design in linear models:

$$\begin{aligned} y_x &= f^T(x)\,\theta + \varepsilon_x\,; \quad x \in \mathcal{X} = \text{design space} \\ E(\varepsilon_x) &= 0, \quad Var(\varepsilon_x) = \sigma^2 \equiv 1 \end{aligned}$$

$x_1, \ldots, x_N$ = exact design
relative frequency$= \xi$ = design

$$\begin{aligned} Var_\xi\left(c^T\hat{\theta}\right) &= \max_{\alpha : M(\xi)\alpha \neq 0} \frac{\left(c^T\alpha\right)^2}{\alpha^T M(\xi)\,\alpha} \\ M(\xi) &= \int_{\mathcal{X}} f(x)\, f^T(x)\, d\xi(x) \end{aligned}$$

The present talk is about two extensions of the c-optimality criterion

$$\xi \to \max_{\alpha : M(\xi)\alpha \neq 0} \frac{\left(c^T\alpha\right)^2}{\alpha^T M(\xi)\,\alpha}$$

to other models.

Extension to parameter-free (infinite-dimensional) lin. models

$\mathcal{H}$ is the set (linear space) of all allowed regression functions defined on the set $\mathcal{X}$ (for example $\mathcal{H} = \left\{\eta(.) = f^T(.)\,\theta,\ \theta \in \mathbb{R}^p\right\}$).
Further,

$$g : \mathcal{H} \rightarrow \mathbb{R}$$

is a linear functional,

$$\|\eta\|_\xi^2 = \int_{\mathcal{X}} \eta^2(x)\, d\xi(x)\ ; \quad \eta \in \mathcal{H}$$

Define

$$\Phi_g(\xi) \equiv \max_{\eta \in \mathcal{H}} \frac{g^2(\eta)}{\|\eta\|_\xi^2}.$$

Remark: THERE IS NO INFORMATION MATRIX IN $\Phi_g(\xi)$.

Extension for the nonlinearly parametrized regression models

$$\begin{aligned} y_x &= \eta(x,\theta) + \varepsilon_x\,; \quad x \in \mathcal{X} \\ E(\varepsilon_x) &= 0, \quad Var(\varepsilon_x) = \sigma^2 \equiv 1 \end{aligned}$$

$h(\theta)$ = (smooth) function to be estimated
$\bar{\theta}$ = the true (guessed) value of θ

Define

$$\begin{aligned} \Psi_h(\xi) &= \min_{\theta \in \mathbb{R}^p, h(\theta) \neq h(\bar{\theta})} \psi_h(\xi;\theta) \\ \text{with} \quad \psi_h(\xi;\theta) &= \frac{\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta\left(x,\bar{\theta}\right)\right]^2 d\xi(x)}{\left[h(\theta) - h\left(\bar{\theta}\right)\right]^2} \end{aligned}$$

Again: THERE IS NO INFORMATION MATRIX IN $\Psi_h(\xi)$.

Basic properties

Property of Extension 1:
If the model is linear and parametrized, and if

$$g\left[f^{T}(.)\,\theta\right] = c^{T}\theta$$

then

$$\Phi_g(\xi) = Var_{\xi}\left(c^{T}\hat{\theta}\right).$$

Property of Extension 2:
If the model is linear and parametrized, and if

$$\begin{aligned} \det\left[M(\xi)\right] &\neq 0, \\ h(\theta) &= c^{T}\theta + d \end{aligned}$$

then

$$\Psi_h(\xi) = \left[Var_{\xi}\left(c^{T}\hat{\theta}\right)\right]^{-1}.$$

Basic properties

Proof:

When

$$\begin{aligned} \eta(x,\alpha) &= f^T(x)\,\alpha, \\ g[\eta(.,\alpha)] &= c^T\alpha, \end{aligned}$$

then

$$\begin{aligned} \alpha^T M(\xi)\,\alpha &= \alpha^T\left[\int_{\mathcal{X}} f(x) f^T(x)\, d\xi(x)\right]\alpha \\ &= \int_{\mathcal{X}} \left[f^T(x)\,\alpha\right]^2 d\xi(x) = \|\eta(.,\alpha)\|_{\xi}^2 . \end{aligned}$$

Hence

$$\max_{\eta\in\mathcal{H}} \frac{g^2(\eta)}{\|\eta\|_{\xi}^2} = \max_{\alpha: M(\xi)\alpha\neq 0} \frac{\left(c^T\alpha\right)^2}{\alpha^T M(\xi)\,\alpha}.$$

Basic properties

Similarly, when

$$\begin{aligned} h(\theta) &= c^T\theta + d, \\ \eta(x,\theta) - \eta(x,\bar{\theta}) &= f^T(x)(\theta - \bar{\theta}), \end{aligned}$$

then

$$\begin{aligned} &\min_{\theta\in\mathbb{R}^p, h(\theta)\neq h(\bar{\theta})} \frac{\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta(x,\bar{\theta})\right]^2 d\xi(x)}{\left[h(\theta) - h(\bar{\theta})\right]^2} = \\ = &\min_{\theta\in\mathbb{R}^p, c^T(\theta-\bar{\theta})\neq 0} \frac{(\theta-\bar{\theta})^T M(\xi)(\theta-\bar{\theta})}{\left[c^T(\theta-\bar{\theta})\right]^2} = \\ = &\left[\max_{\alpha\neq 0} \frac{(c^T\alpha)^2}{\alpha^T M(\xi)\alpha}\right]^{-1}. \end{aligned}$$

□

The interpretation of $\Phi_g(\xi)$

Take

$$x_1, \ldots, x_N = \text{exact design},$$

$$\xi = \text{“frequency” design.}$$

Define the **observed random measure**

$$Y_\xi(B) \equiv \frac{1}{N} \sum_{i=1, x_i \in B}^{N} y_{x_i}.$$

Then for any $\eta \in \mathcal{H}$

$$\begin{aligned} E_\eta(Y_\xi(B)) &= \int_B \eta(x)\, d\xi(x), \\ Cov\left[Y_\xi(B_1), Y_\xi(B_2)\right] &= \xi(B_1 \cap B_2). \end{aligned}$$

A linear estimator is an integral with respect to this observed random measure.

The interpretation of $\Phi_g(\xi)$

Theorem (cf. A. Pázman, Kybernetika 1978)

The functional g is linearly estimable under ξ iff there is a function

$$l \in L_2(\xi)$$

such that

$$g(\eta) = \int_{\mathcal{X}} l(x)\,\eta(x)\,d\xi(x)\,; \quad \eta \in \mathcal{H}.$$

Then

$$Var_{\xi}(BLUE) = \max_{\eta \in \mathcal{H}} \frac{g^2(\eta)}{\|\eta\|_{\xi}^2} = \|l_{\xi}\|_{\xi}^2\,,$$

where l_{ξ} = the $L_2(\xi)$-orthogonal projection of l onto $\mathcal{H}$.

The interpretation of $\Psi_h(\xi)$

If $\left\|\theta - \bar{\theta}\right\|$ is small, then

$$
\begin{aligned}
\Psi_h(\xi) &= \min_\theta \frac{\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta\left(x,\bar{\theta}\right)\right]^2 d\xi(x)}{\left[h(\theta) - h\left(\bar{\theta}\right)\right]^2} \\
&= \min_\theta \frac{\int_{\mathcal{X}} \left[\frac{\partial \eta(x,\theta)}{\partial \theta}\,|_{\bar{\theta}}\,\left(\theta - \bar{\theta}\right) + o\left(\left\|\theta - \bar{\theta}\right\|\right)\right]^2 d\xi(x)}{\left[\frac{\partial h(\theta)}{\partial \theta}\,|_{\bar{\theta}}\,\left(\theta - \bar{\theta}\right) + o\left(\left\|\theta - \bar{\theta}\right\|\right)\right]^2} \\
&= \min_\theta \frac{\int_{\mathcal{X}} \left(\theta - \bar{\theta}\right)^T M\left(\xi,\bar{\theta}\right)\left(\theta - \bar{\theta}\right)}{\left[\frac{\partial h(\theta)}{\partial \theta}\,|_{\bar{\theta}}\,\left(\theta - \bar{\theta}\right)\right]^2} + o\left(\left\|\theta - \bar{\theta}\right\|\right) \\
&= \left\{\left[\frac{\partial h(\theta)}{\partial \theta^T}\right]_{\bar{\theta}} M^{-1}\left(\xi,\bar{\theta}\right)\left[\frac{\partial h(\theta)}{\partial \theta}\right]_{\bar{\theta}}\right\}^{-1} + o\left(\left\|\theta - \bar{\theta}\right\|\right)
\end{aligned}
$$

$\Rightarrow$ standard local c-optimality criterion in nonlinear models.

The interpretation of $\Psi_h(\xi)$

However, there is much more in $\Psi_h(\xi)$, namely

$$\Psi_h(\xi) = \min_{\theta \in \mathbb{R}^p, h(\theta) \neq h(\bar{\theta})} \frac{\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta\left(x,\bar{\theta}\right)\right]^2 d\xi(x)}{\left[h(\theta) - h\left(\bar{\theta}\right)\right]^2}$$

detects the lack of identifiability:

1) If

$$\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta\left(x,\bar{\theta}\right)\right]^2 d\xi(x) = 0$$

$$\Downarrow$$

we miss a standard condition for consistency and asymptotical normality of $\hat{\theta}$.

2)

If $\int_{\mathcal{X}} \left[\eta(x,\theta) - \eta(x,\bar{\theta})\right]^2 d\xi(x)$ much smaller than $\left[h(\theta) - h\left(\bar{\theta}\right)\right]^2$

$$\Downarrow$$

the estimator $h(\hat{\theta})$ can not distinguish $h\left(\bar{\theta}\right)$ from $h(\theta)$.

Theorem

The function $\Phi_g(\xi)$ is convex in ξ.
The function $\Psi_h(\xi)$ is concave in ξ.

Proof:
The concavity of

$$\Psi_h(\xi) = \min_{\theta \in \mathbb{R}^p, h(\theta) \neq h(\bar{\theta})} \psi_h(\xi; \theta)$$

is evident, since

$$\psi_h(\xi; \theta) = \frac{\int_{\mathcal{X}} \left[\eta(x, \theta) - \eta(x, \bar{\theta})\right]^2 d\xi(x)}{\left[h(\theta) - h(\bar{\theta})\right]^2}$$

is linear in ξ.

Further, for every $\eta \in \mathcal{H}$

$$0 \leq \|I_\xi - \eta\|_\xi^2 = \|I_\xi\|_\xi^2 - 2g(\eta) + \|\eta\|_\xi^2 .$$

Hence

$$\Phi_g(\xi) = \|I_\xi\|_\xi^2 \geq 2g(\eta) - \|\eta\|_\xi^2$$

and the equality is attained if $\eta = I_\xi$.

$$\Phi_g(\xi) = \max_{\eta \in \mathcal{H}} \left\{ 2g(\eta) - \int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right\}$$

which is evidently convex in ξ.

□

Convexity, concavity and directional derivatives

A convex (concave) function has always a **directional derivative** (at the point ξ and in a direction given by δ)

$$\nabla_{\delta}\Phi_g(\xi) = \lim_{\beta \to 0} \frac{\Phi_g[(1-\beta)\xi + \beta\delta] - \Phi_g(\xi)}{\beta}$$

and similarly for $\Psi_h(\xi)$.

As well known, a design μ is Φ_g-optimal if and only if

$$\nabla_{\delta}\Phi_g(\mu) \geq 0, \quad \forall\delta$$

Similarly for the Ψ_h-optimality we require

$$\nabla_{\delta}\Psi_h(\mu) \leq 0, \quad \forall\delta$$

Theorem

We have

$$\begin{aligned}\nabla_\delta \Phi_g(\xi) &= \int_{\mathcal{X}} l_\xi^2(x)\, d\xi(x) - \int_{\mathcal{X}} l_\xi^2(x)\, d\delta(x),\\ \nabla_\delta \Psi_h(\xi) &= \min_{\theta\in\Theta(\xi)} \psi_h(\delta,\theta) - \Psi_h(\xi),\\ \Theta(\xi,h) &= \arg\min_{\theta\in\Theta} \psi_h(\xi,\theta).\end{aligned}$$

The first expression requires to compute a projection in $L_2(\xi)$, the second requires to solve a minimization problem to obtain $\Theta(\xi,h)$.

Proof: From books on minimax solutions we obtain

$$\begin{aligned}\nabla_\delta\left[\min_{\theta\in\Theta}\psi_h(\xi,\theta)\right] &= \min_{\theta\in\Theta(\xi,h)}\left[\nabla_\delta\psi_h(\xi,\theta)\right]\\ &= \min_{\theta\in\Theta(\xi,h)}\frac{d}{d\beta}\left[\psi_h((1-\beta)\xi(x)+\beta\delta,\theta)\right]\\ &= \min_{\theta\in\Theta(\xi,h)}\left[\psi_h(\delta,\theta)-\psi_h(\xi,\theta)\right]\end{aligned}$$

since $\psi_h(\xi,\theta)$ is linear in ξ.

Similarly, from

$$\Phi_g(\xi) = \max_{\eta \in \mathcal{H}} \left\{ 2g(\eta) - \int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right\}$$

we have

$$\begin{aligned} & \nabla_\delta \left[\max_{\eta \in \mathcal{H}} \left\{ 2g(\eta) - \int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right\} \right] \\ = \quad & \max_{\eta \in \mathcal{H}(\xi)} \left[\nabla_\delta \left(- \int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right) \right] \end{aligned}$$

where

$$\begin{aligned} \mathcal{H}(\xi) \quad &= \quad \arg\max_{\eta \in \mathcal{H}} \left\{ 2g(\eta) - \int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right\} = \{ I_\xi \}, \\ \nabla_\delta \left(\int_{\mathcal{X}} \eta^2(x)\, d\xi(x) \right) \quad &= \quad \int_{\mathcal{X}} \eta^2(x)\, d\delta(x) - \int_{\mathcal{X}} \eta^2(x)\, d\xi(x). \end{aligned}$$

□

Theorem (Equivalence for Φ_g in parameter-free models)

A design μ is Φ_g optimal if and only if

$$\max_{x\in\mathcal{X}} I_{\mu}^{2}(x) = \int_{\mathcal{X}} I_{\mu}^{2}(x)\, d\mu(x)$$

Proof:

$$\nabla_{\delta}\Phi_g(\xi) = \int_{\mathcal{X}} I_{\xi}^{2}(x)\, d\xi(x) - \int_{\mathcal{X}} I_{\xi}^{2}(x)\, d\delta(x) \geq 0; \quad \forall\delta$$

□

Particular case: linear model with $\det M(\mu) \neq 0$

$$\begin{aligned} I_{\mu}(x) &= c^T M^{-1}(\mu) f(x) \\ &\Downarrow \\ \max_{x\in\mathcal{X}} \left[c^T M^{-1}(\mu) f(x)\right]^2 &= c^T M^{-1}(\mu)\, c \end{aligned}$$

Theorem (Equivalence for Ψ_h in nonlinear models)

Suppose that $\Theta(\mu, h) = \{\theta_{\mu,h}\}$ = a unique point.
Then a design μ is Ψ_h-optimal if and only if

$$\max_{x \in \mathcal{X}} \left[\eta(x, \theta_{\mu,h}) - \eta\left(x, \bar{\theta}\right)\right]^2 = \int_{\mathcal{X}} \left[\eta(x, \theta_{\mu,h}) - \eta\left(x, \bar{\theta}\right)\right]^2 d\mu(x)$$

Proof:

$$\forall \delta \quad \nabla_\delta \Psi_h(\mu) = \psi_h(\delta, \theta_{\mu,h}) - \psi_h(\xi, \theta_{\mu,h}) \leq 0$$

$$\Downarrow$$

$$\forall \delta \qquad \frac{\int_{\mathcal{X}} \left[\eta(x, \theta_{\mu,h}) - \eta\left(x, \bar{\theta}\right)\right]^2 d\delta(x)}{\left[h(\theta_{\mu,h}) - h\left(\bar{\theta}\right)\right]^2} \leq$$

$$\leq \Psi_h(\mu) = \frac{\int_{\mathcal{X}} \left[\eta(x, \theta_{\mu,h}) - \eta\left(x, \bar{\theta}\right)\right]^2 d\mu(x)}{\left[h(\theta_{\mu,h}) - h\left(\bar{\theta}\right)\right]^2}$$

□

Particular case: linear model with $\det M(\mu) \neq 0$

$$\begin{aligned} \theta_{\mu,h} - \bar{\theta} &= M^{-1}(\mu)\, c \\ \eta(x, \theta_{\mu,h}) - \eta(x, \bar{\theta}) &= f^T(x)\, M^{-1}(\mu)\, c \\ &\Downarrow \\ \max_{x \in \mathcal{X}} \left[f^T(x)\, M^{-1}(\mu)\, c \right]^2 &= \int_{\mathcal{X}} \left[f^T(x)\, M^{-1}(\mu)\, c \right]^2 d\mu(x) \\ &= c^T M^{-1}(\mu)\, c \end{aligned}$$

Remark: If $\Theta(\mu, h)$ has finitely many points we have:

Theorem

A design μ is Ψ_h optimal if and only if there is a probability measure λ on $\Theta(\mu, h)$ such that $\sum_{\theta \in \Theta(\mu,h)} \lambda(\theta) [\psi_h(\delta, \theta) - \psi_h(\mu, \theta)] \leq 0 \,; \quad \forall \delta,$ or, equivalently,

$$\max_{x \in \mathcal{X}} \sum_{\theta \in \Theta(\mu,h)} \lambda(\theta) \left[\eta(x, \theta) - \eta(x, \bar{\theta}) \right]^2 = \Psi_h(\mu)\,.$$

Extending Elfvings's theorem

Theorem (Elfving's theorem (1952) in parametrized linear models)

In the linear model $\eta(x,\theta) = f^T(x)\theta$ *let*

$$\mathcal{S} \equiv co\left\{f\left(\mathcal{X}\right) \cup \left[-f\left(\mathcal{X}\right)\right]\right\}.$$

Then

$$\mu \in \arg\min_{\xi} Var_{\xi}\left(c^T\hat{\theta}\right)$$

if and only if $\exists \beta \in \mathbb{R}$ *such that*

i) $\beta c \in$ *boundary of* $\mathcal{S}$

ii) $\beta c = \int_{\mathcal{X}} \delta\left(x\right) f\left(x\right) d\mu\left(x\right)$ *for some* $\delta\left(x\right) \in \{-1, 1\}$

Extensions to other criteria by Studden, Dette etc.
Numerical computation:
- J. López Fidalgo & Rodrígues-Díaz (Metrika)
- R. Harman & T. Jurík (Comp. Stat. & Data Analysis, 2008): a linear programming method

Extension of Elfving's theorem to Φ_g in parameter-free models

Theorem (cf. A. Pázman, Kybernetika 1978)

Hahn-Banach theorem $\Rightarrow$ *the functional g can be extended from $\mathcal{H}$ to the set $C(\mathcal{X})$ so that*

$$\max_{\eta \in \mathcal{H}} \frac{[g(\eta)]^2}{\|\eta\|_\infty^2} = \max_{f \in \mathcal{C}(\mathcal{X})} \frac{[g(f)]^2}{\|f\|_\infty^2}.$$

Riesz theorem $\Rightarrow$ *there is a generalized measure ν on $\mathcal{X}$ such that*

$$g(f) = \int_{\mathcal{X}} f(x)\, d\nu(x) \quad ; f \in C(\mathcal{X}).$$

Let $\nu = \nu^+ - \nu^-$ be the Jordan decomposition of ν.

Then the design $\mu = \dfrac{\nu^+ + \nu^-}{\nu^+(\mathcal{X}) + \nu^-(\mathcal{X})}$ is Φ_g-optimal.

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

No question is so difficult to answer as that to which the answer is obvious.

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

No question is so difficult to answer as that to which the answer is obvious.

Everything happens to everybody sooner or later if there is time enough.

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

No question is so difficult to answer as that to which the answer is obvious.

Everything happens to everybody sooner or later if there is time enough.

Science . . . never solves a problem without creating ten more.

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

No question is so difficult to answer as that to which the answer is obvious.

Everything happens to everybody sooner or later if there is time enough.

Science . . . never solves a problem without creating ten more.

It is the mark of a truly intelligent person to be moved by statistics.

Some words from G.B. Shaw before the today's evening to which you are all cordially invited

No question is so difficult to answer as that to which the answer is obvious.

Everything happens to everybody sooner or later if there is time enough.

Science . . . never solves a problem without creating ten more.

It is the mark of a truly intelligent person to be moved by statistics.

Imagination is the beginning of creation. You imagine what you desire, you will what you imagine and at last you create what you will.