Measures everywhere

Design of experiments

Sergei Zuyev

University of Strathclyde, Glasgow, U.K.

Optimal design of experiments

Regression model:

$$y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + \sigma dw(x) = f(x)\beta + dw(x) \qquad (1)$$

- $x \in X$ – *design space*,
- $\beta = (\beta_1, \ldots, \beta_k)^\top$ – unknown parameters,
- $f(x) = (f_1(x), \ldots, f_k(x))$ – a row of linearly independent on X functions, and
- $\sigma dw(x)$ is an orthogonal stationary white noise with variance σ^2.

Given n observations $x_1, \ldots, x_n$, the Least Square Estimator $\widehat{\beta} = \widehat{\beta}(x_1, \ldots, x_n)$ minimises

$$\sum_{i=1}^{n} \Big(y(x_i) - \sum_{j=1}^{k} \beta_j f_j(x_i) \Big)^2 . \qquad (2)$$

How to choose observation points $x_1, \ldots, x_n$ to achieve better properties of $\widehat{\beta}$?

- We have a minimisation problem of a function of n variables: can be solved directly if n is not large.

- When n *is* large, the design is represented by a design measure $\mu(dx)$: a probability distribution on X describing the frequency of taking x as an observation point.

D-optimal designs

D-optimal design minimises the *generalised variance*:

$$\det \| \mathbf{cov}(\widehat{\beta}_i, \widehat{\beta}_j) \|$$

If $\mu(dx)$ is a discrete design measure, the generalised variance equals $\sigma^2 \det M^{-1}(\mu)$, where

$$M(\mu) = \int f(x)^\top f(x)\, \mu(dx)$$

is the *information matrix*.

General Equivalence Theorem

Theorem 1. *(Kiefer-Wolfowitz)*

- *A D-optimal design measure can be chosen to have at most k atoms.*
- *A measure μ provides a D-optimal design if and only if the function*

$$d(x, \mu) = f(x)M^{-1}(\mu)f(x)^\top$$

called the standardised variance of the predicted response at point x*, achieves its maxima at the atoms of* μ*.*

Ψ-optimal designs

Ψ-optimal design minimises $\Psi(M(\mu))$ for a given differentiable function $\Psi : \mathbb{R}^{k^2} \mapsto \mathbb{R}$.

NB. For D-optimal design, $\Psi(M(\mu)) = -\log\det M(\mu)$.

Usual approach: optimal μ is sought in 2 steps:

$$\Psi(M(\mu)) \quad \overset{\text{Step I}}{\longrightarrow} \quad M(\mu) \quad \overset{\text{Step II}}{\longrightarrow} \quad \mu\,.$$

- Step I – optimisation in $\mathbb{R}^{(k+1)/2}$;
- Step II – identify μ by its information matrix $M(\mu)$.

Direct approach: optimisation on the space of positive measures $\mathbb{M}_+$:

$$\psi(\mu) = \Psi(M(\mu)) \longrightarrow \min \quad \text{subject to} \quad \mu(X) = 1$$

Generalised Kiefer-Wolfowitz theorem

Theorem 2. *Let* $m_{ij} = \int f_i(x) f_j(x)\, \mu(dx)$ *be* (i, j)*-th entry of the information matrix* $M(\mu)$*. Then the* Ψ*-optimal design measure* μ *satisfies*

$$\begin{cases} f(x) D\Psi(M)(\mu) f^{\top}(x) = u & \mu - \text{a.e.}\,, \\ f(x) D\Psi(M)(\mu) f^{\top}(x) \geq u & \forall x \in X\,, \end{cases} \tag{3}$$

where

$$D\Psi(M)(\mu) = \left\| \frac{\partial \Psi(M)}{\partial m_{ij}}(\mu) \right\|_{ij} .$$

Proof

We just need to evaluate the gradient of Ψ and apply our first-order necessary condition. By the chain rule,

$$
\begin{aligned}
D\Psi(M(\mu))[\eta] &= \sum_{i,j} \frac{\partial \Psi(M)}{\partial m_{ij}}(\mu) D m_{ij}(\mu)[\eta] \\
&= \sum_{i,j} \frac{\partial \Psi(M)}{\partial m_{ij}}(\mu) \int f_i(x) f_j(x) \eta(dx) \\
&= \int f(x) D\Psi(M)(\mu) f^\top(x) \eta(dx)
\end{aligned}
$$

so that the gradient function is $f(x) D\Psi(M)(\mu) f^\top(x)$.

Additional constraints

Assume that we now impose additional constraints on the design measure, e. g.

$$\begin{cases} H_i(\mu) = 0\,, & i = 1, \ldots, l\,; \\ H_i(\mu) \le 0\,, & i = l+1, \ldots, m\,. \end{cases} \tag{4}$$

where H_i are Fréchet differentiable functions with gradients $h_i(x, \mu)$.

As before, this would only change the RHS in (3): u is replaced by the linear combination $\sum_{i=1}^{m} u_i h_i(x, \mu)$. This would be impossible to guess using a common two-step method!

Constrained Kiefer-Wolfowitz theorem

Theorem 3. *If μ is a regular Ψ-optimal measure under constraints (4), then there exist Lagrange multipliers $u_1, \ldots, u_m$ with $u_j \leq 0$ if $H_j(\mu) = 0$ and $u_j = 0$ if $H_j(\mu) < 0$ for $j \in \{l+1, \ldots, m\}$, such that*

$$\begin{cases} f(x) D\Psi(M)(\mu) f^\top(x) = \sum_{i=1}^m u_i h_i(x, \mu) & \mu - \text{a.e.}\,, \\ f(x) D\Psi(M)(\mu) f^\top(x) \geq \sum_{i=1}^m u_i h_i(x, \mu) & \forall x \in X\,. \end{cases} \quad (5)$$

Kiefer-Wolfowitz vs. General Equivalence

The Kiefer-Wolfowitz theorem 1 is also called General Equivalence Theorem as it provides criterium for μ to be D-optimal. As, in general, Ψ-function may not have a unique minimum, the 'if and only if' statement is no longer possible.

- **But:** if Ψ is convex (as for D-optimal designs, where $\Psi(M(\mu)) = -\log\det M(\mu)$), then (5) also becomes necessary and sufficient condition for the Ψ-optimality.

References

- J. Kiefer and J. Wolfowitz. Equivalence of two extremal problems, Canad. J. Math., **14**, 1960, 849–879.
- I. Molchanov and S. Zuyev. Optimisation in space of measures and optimal design. ESAIM (to appear).
- A.C. Atkinson and A.N. Donev. Optimum Experimental Designs, Oxford, 1992.

❏ http://www.stams.strath.ac.uk/~sergei