Optimal design for models of computer experiments

Andrey Pepelyshev

The statement of problem

Consider the model

$$y_j = y_j(t_j) = \theta_1 f_1(t_j) + \ldots + \theta_m f_m(t_j) + \varepsilon_j$$

where $t_j \in [-T, T]$, $j = 1, ..., N$ and $\mathbf{E}\varepsilon_j\varepsilon_i = K(t_i, t_j) = \sigma^2\rho(t_j - t_i)$.
For the estimate

$$\hat{\theta}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$$

the exact design problem has the form

$$\text{Var}\left(\hat{\theta}_{\text{OLS}}\right) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Sigma X(\mathbf{X}^T\mathbf{X})^{-1} \overset{\Phi\text{ crit}}{\longrightarrow} \min_{t_1,\ldots,t_N}.$$

NB: We can use the weighted least squares estimate in practice but here we study designs that are optimal for the ordinary least squares estimate by two reasons. First, the OLS-optimal design is more easy to construct. Second, the OLS-optimal design is very efficient for WLS estimation.

Asymptotic settings

Let the design points $\{t_1, \ldots, t_N\}$ be generated by the quantiles of a distribution function,

$$t_{iN} = a\left((i-1)/(N-1)\right),\ i = 1, \ldots, N,$$

where the function $a : [0,1] \to [-T, T]$ is the inverse of a distribution function.
Let ξ be a design measure corresponding to $a(\cdot)$.
Under asymptotic settings, the design problem has the form

$$D(\xi) = M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi) \overset{\Phi\text{ crit}}{\longrightarrow} \min_{\xi}$$

where $M(\xi) = \int f(u)f^T(u)\xi(du)$ and $B(\xi,\nu) = \int\int \rho(u-v)f(u)f^T(v)\xi(du)\nu(dv)$.

The optimality condition

Define $C = \frac{\partial\Phi(D)}{\partial D} = \left(\frac{\partial\Phi(D)}{\partial D_{ij}}\right)_{i,j=1,\ldots,m}$,

$$\varphi(x,\xi) = f^T(x)M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)\,C(\xi)M^{-1}(\xi)f(x),$$

$$b(x,\xi) = f^T(x)B^{-1}(\xi,\xi)\int \rho(u-x)f(u)\xi(u),$$

Theorem 1. Let ξ^* be any design minimizing the functional $\Phi(D(\xi))$. Then the inequality

$$\varphi(x,\xi^*) \le b(x,\xi^*)$$

holds for all $x \in \mathcal{X}$. Moreover, there is equality for ξ^*-almost all x.

Theorem 2. Let ξ^* be any D-optimal design. Then for all $x \in \mathcal{X}$ we have

$$d(x,\xi^*) \le b(x,\xi^*)$$

where the functions d is defined by

$$d(x,\xi) = f^T(x)M^{-1}(\xi)f(x).$$

Moreover, there is equality for ξ^*-almost all x.

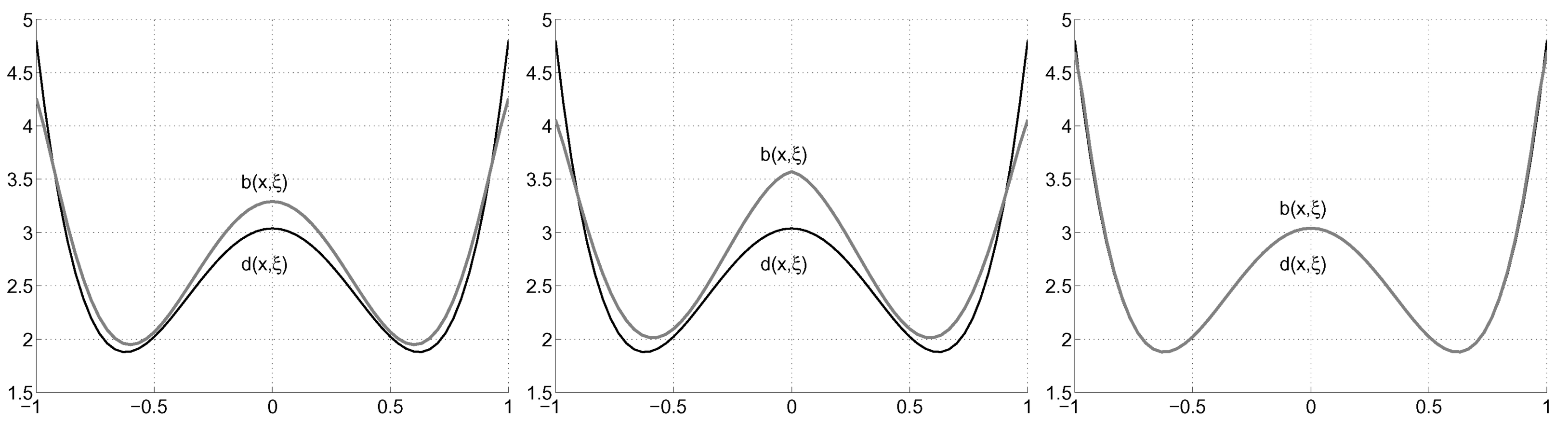

Figure 1. The functions $b(x,\xi)$ and $d(x,\xi)$ for the regression model with $f(x) = (1, x, x^2)^T$ and the covariance kernels $\rho(u-v) = e^{-|u-v|}$ (left), $\rho(u-v) = \max(0, 1-|u-v|)$ (middle) and $\rho(u-v) = -\log(u-v)^2$ (right), and the arcsine design ξ.

The polynomial regression model and $\rho(x) \propto \ln x^2$

Theorem 3. Consider the polynomial regression model with $f(x) = (1, x, x^2, \ldots, x^{m-1})^T$, $x \in [-1,1]$, and the covariance function is $\rho(x) = \gamma - \beta\ln x^2$ with $\gamma \ge 0$, $\beta > 0$. Then the design with the arcsine density satisfies the necessary conditions for universal optimality.

Sketch of proof. It can be verified that the (generalized) arcsine design satisfies the optimality condition for the c-criterion for all c.

The Mercer theorem

Define a linear operator T_K by

$$[T_K\varphi](x) = \int_a^b K(x,s)\varphi(s)\,ds.$$

Theorem. (Mercer, 1909) Suppose $K(s,t)$ is a continuous symmetric non-negative definite kernel. Then there is an orthonormal basis $\{\varphi_i\}_i$ of $L_2[a,b]$ consisting of eigenfunctions of T_K such that the corresponding sequence of eigenvalues $\{\lambda_i\}_i$ is nonnegative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $\mathcal{X} = [a,b]$ and K has the representation

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j\,\varphi_j(s)\,\varphi_j(t)$$

where the convergence is absolute and uniform.

Examples.
1. Let $K(x,u) = e^{-\lambda|x-u|}$ and $\mathcal{X} = [-1,1]$. Then $\varphi_k(x) = \sin(\omega_k x + k\pi/2)$, where $\omega_1, \omega_2, \ldots$ are positive roots of the equation $\tan(2\omega) = -2\lambda\omega/(\lambda^2 - \omega^2)$.
2. Let $K(x,u) = \min\{x,u\}$ and $\mathcal{X} = [0,1]$. Then $\varphi_k(x) = \sin((k+1/2)\pi x)$.
3. Let $K(s,t) = \rho(s-t)$ and $\rho(t) = \rho(t-1)$, $\mathcal{X} = [0,1]$. Then $\varphi_j(x) = \sqrt{2}\cos(2\pi(j-1)x)$ for $j \ge 1$ and $\varphi_0(x) = 1$.

Optimality of the uniform design in special cases

Theorem 3. Let $\mathcal{X}$ be a finite interval and $K(x,u) = \sum_k \lambda_k\varphi_k(x)\varphi_k(u)$ be the expansion from the Mercer theorem. Consider the regression model with $f(x) = (\varphi_{i_1}(x), \ldots, \varphi_{i_m}(x))^T$, $i_j \ne i_l$, and the covariance kernel $K(x,u)$. Then the design with uniform density satisfies the necessary conditions for universal optimality.

Sketch of proof. Note that $M(\xi)$ and $B(\xi,\xi)$ are diagonal matrices. Then it is easy to verify that the optimality condition of Theorem 1 is fulfilled for the c-criterion for all c.

Efficiencies of the uniform and arcsine designs

Let us study the D-efficiency of the uniform design and the arcsine design for the linear regression model with $f(x) = (1, x, \ldots, x^{m-1})^T$ and different correlation functions.

We determine the D-efficiency as

$$\text{Eff}(\xi) = \left(\frac{\det D(\xi^*)}{\det D(\xi)}\right)^{1/m}.$$

In our numerical study we computed the D-optimal design ξ^* by the multiplicative algorithm. In Tables 1–3 we can observe that the efficiency of the arcsine design is mainly larger than the efficiency of the uniform design. Moreover, the difference between efficiencies of the arcsine design and the uniform design increases as m increases. In addition, the efficiency of the uniform design and the arcsine design decreases as m increases.

Table 1. D-Efficiencies of the uniform design ξ_u and the arcsine design ξ_a for the model with $f(x) = (1, x, \ldots, x^{m-1})^T$ and the gaussian correlation function $\rho(x) = e^{-\lambda x^2}$.

	λ	0.5	1.5	2.5	3.5	4.5	5.5
$m=1$	Eff(ξ_u)	0.758	0.789	0.811	0.830	0.842	0.853
	Eff(ξ_a)	0.841	0.907	0.924	0.932	0.934	0.935
$m=2$	Eff(ξ_u)	0.756	0.698	0.709	0.725	0.739	0.753
	Eff(ξ_a)	0.843	0.833	0.853	0.868	0.877	0.885
$m=3$	Eff(ξ_u)	0.803	0.662	0.684	0.699	0.711	0.720
	Eff(ξ_a)	0.866	0.771	0.818	0.844	0.859	0.869
$m=4$	Eff(ξ_u)	0.797	0.630	0.617	0.627	0.648	0.665
	Eff(ξ_a)	0.842	0.713	0.722	0.746	0.776	0.799

Table 2. D-Efficiencies of the uniform design ξ_u and the arcsine design ξ_a for the model with $f(x) = (1, x, \ldots, x^{m-1})^T$ and the exponential correlation function $\rho(x) = e^{-\lambda|x|}$.

	λ	0.5	1.5	2.5	3.5	4.5	5.5
$m=1$	Eff(ξ_u)	0.913	0.888	0.903	0.919	0.933	0.944
	Eff(ξ_a)	0.966	0.979	0.987	0.980	0.968	0.954
$m=2$	Eff(ξ_u)	0.857	0.832	0.847	0.867	0.886	0.901
	Eff(ξ_a)	0.942	0.954	0.970	0.975	0.973	0.966
$m=3$	Eff(ξ_u)	0.832	0.816	0.826	0.842	0.860	0.876
	Eff(ξ_a)	0.934	0.938	0.954	0.968	0.976	0.981
$m=4$	Eff(ξ_u)	0.826	0.818	0.823	0.835	0.849	0.864
	Eff(ξ_a)	0.934	0.936	0.945	0.957	0.967	0.975

References

H. Dette, A. Pepelyshev, A. Zhigljavsky (2011) Optimal design for linear models with correlated observations. in progress.