Minimum bias designs under random contamination: application to polynomial spline models

Dave Woods
University of Southampton, UK

D.C.Woods@maths.soton.ac.uk

Mean squared error

- Assumed model for an observation

$$Y = f(x) + \varepsilon$$

- True model

$$Y = f(x) + \varphi(x) + \varepsilon$$

- Aim is to estimate $f(x)$

- Average mean squared error (AMSE)
  - error in predictions over design region
  - approximated over a discrete grid of r points

$$\frac{n}{r\sigma^2} \sum_{i=1}^{r} E\left\{ \hat{f}(x_i) - f(x_i) - \varphi(x_i) \right\}^2$$

Variance and bias

For linear model $f(x_i) = f_i^T \beta$

- AMSE = Variance, V + Squared Bias, B

$$V = \frac{n}{r} tr \left\{ F(X^T X)^{-1} F^T \right\} \qquad B = \frac{n}{r\,\sigma^2} \varphi^T P^T P \varphi$$

where $\varphi = \left(\varphi(x_1) \ldots \varphi(x_r)\right)^T$

and

$$P = F(X^T X)^{-1} X^T D - I$$

- a known form is often assumed for $\varphi(x)$ e.g. Box & Draper (1959), Montepiedra & Fedorov (1997)

Random contamination

- Assume $\varphi(x)$ is a realisation of a random variable $\Phi(x)$

- Population of true models

$$Y = f(x) + \Phi(x) + \varepsilon$$

- Random contamination implies random bias for given assumed model and design

- Notz (1989) and Allen *et al* (2003) also assumed random contamination as known specified higher order polynomial terms with random coefficients

Design selection criteria based on bias

- Minimise expected bias ("EB-optimal")

$$E(B) = \frac{n}{r\sigma^2} tr\left\{\mathrm{P}^{\mathrm{T}}\mathrm{P}E\left[\Phi\Phi^{\mathrm{T}}\right]\right\}$$

- Minimise variance bias ("VB-optimal")

$$V(B) = \frac{n^2}{r^2\sigma^4} V\left(tr\left\{\mathrm{P}^{\mathrm{T}}\mathrm{P}\Phi\Phi^{\mathrm{T}}\right\}\right)$$

- Minimise percentile bias ("PB-optimal")
  - find the design that minimises b >0 such that

$$P(B < b) = p$$

for 0< p ≤ 1

Implementation

- Mathematically intractable for even simple cases

- Modified Fedorov exchange algorithm

- Embedded Monte Carlo simulation to approximate properties of bias distribution

- EB-optimality is computationally efficient – each design search only requires one approximation of $E\left[\Phi\Phi^{T}\right]$

Polynomial splines

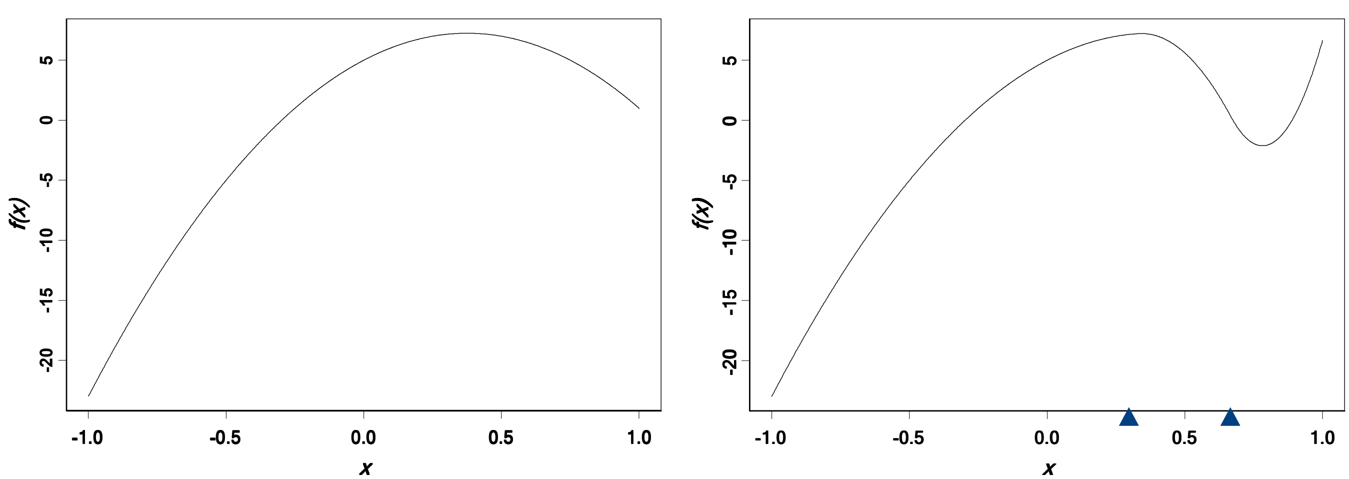
- Allow different low degree polynomials on different sections - separated by l knots τ_j

- For one factor, assumed model

$$f(x) = \sum_{i=0}^{d} \beta_i x^i + \sum_{j=1}^{l} \beta_{d+j} (x - \tau_j)_+^d$$

- truncated power basis

- knot locations τ_j are known
- but uncertainty about additional knots



Spline contamination

- Contamination $\Phi(x)$ has the form

$$\Phi(x) = \sum_{i=1}^{K} \Gamma_i (x - \Lambda_i)_+^d$$

K, Λ_i and Γ_i are random variables
- prior distributions

Example

- $n = 4$ design points

Assumed model

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

Spline contamination with

– $K \sim$ Poisson(μ_k)
– $\Lambda_i \sim$ Uniform(l_1, l_2)
– $\Gamma_i \sim$ N(μ_p, σ_p^2)

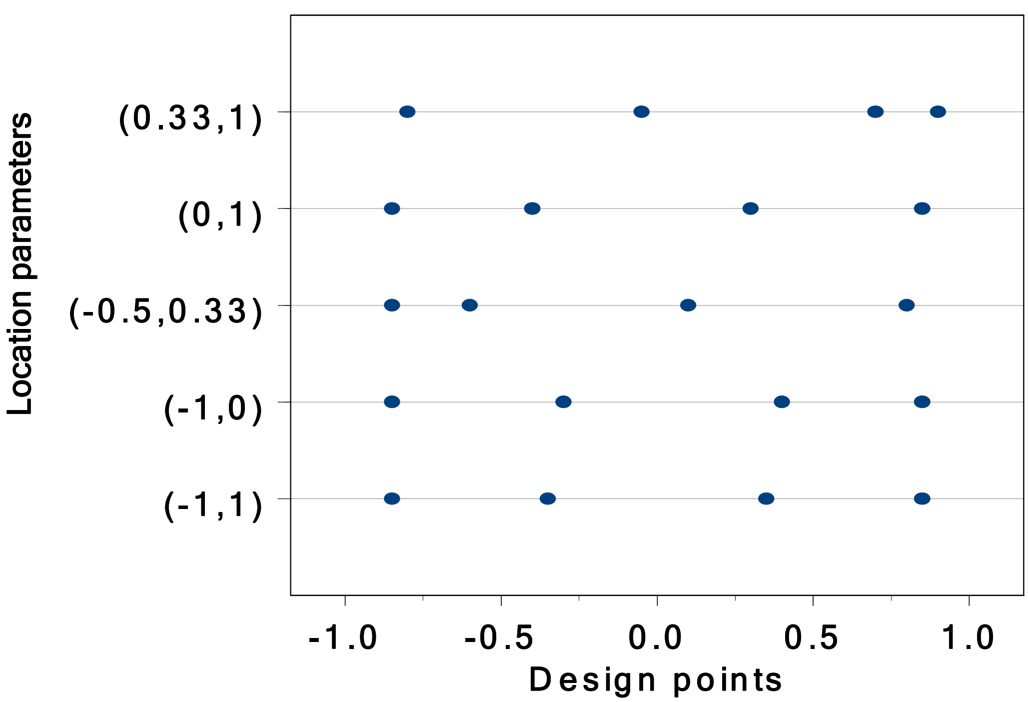
Example

- Varying μ_k, μ_p, σ_p^2

μ_k, μ_p, σ_p^2	EB-optimal design	EB evaluated at			
		2,0,1	2,10,100	15,0,100	15,10,1
2,0,1	I	0.01	3.26	7.91	72.99
2,10,100	II	0.01	3.39	8.09	74.97
15,0,100	I	0.01	3.26	7.91	72.99
15,10,1	I	0.01	3.26	7.91	72.99

I = {-0.85, -0.35, 0.35, 0.85} II = {-0.85, -0.55, 0.15, 0.8}

- Varying l_1, l_2



Comparison of designs using variance and percentile bias

- Designs found with parameters

$\mu_k=2$ $\mu_p=10$ $\sigma_p^2=1$ $l_1=0$ $l_2=1/3$

Criterion	Design	EB	VB	P=0.95
EB	-0.85, -0.55, 0.15, 0.8	6.97	76.8	23.8
VB	-0.75, -0.05, 0.65, 0.85	6.79	76.5	25.0
PB P=0.95	-0.85, -0.45, 0.2, 0.8	6.96	77.5	25.0

Findings from studies

Results from a range of empirical studies agree with the example

- EB-optimal designs appear robust to the values of μ_k, μ_p, σ_p^2....
- ....but not to the values of l_1 and l_2
- The size of the expected bias depends most on μ_p and l_1, l_2
- EB-optimal designs perform well under the other bias criteria

Conclusions and future research

EB-optimal designs

- have more support points than designs from variance based criteria
- are efficient under other random bias criteria
- are computationally practical

Ideas extend to an expected AMSE criterion

Future work

- application to models in laser chemistry

  - supported by EPSRC Combechem E-science grant