Some problems in robustness of design

Doug Wiens

Statistics Centre

Dept. of Mathematical and Statistical Sciences

University of Alberta

- Why robustness of design?
  - The models for which we construct designs are generally at best approximations.
  - The 'best' design for a slightly wrong model can be much more than slightly sub-optimal.
  - Although we will fit the assumed, 'ideal' model, we should design for protection against biases arising from any of a range ('neighbourhood') of alternate models.
- Illustrations to follow are based on simple straight line regression (SLR) – even here there are intriguing, and frustratingly difficult, open problems.
- More complex situations discussed later, as time allows.

- SLR: We fit the model

$$E\left[Y\left(x\right)\right]=\boldsymbol{f}'\left(x\right)\boldsymbol{\theta}=\theta_0+\theta_1 x,$$

(i.e. $\boldsymbol{f}'\left(x\right)=\left(1,x\right)$) for $x\in\left[-1,1\right]$ but are concerned that

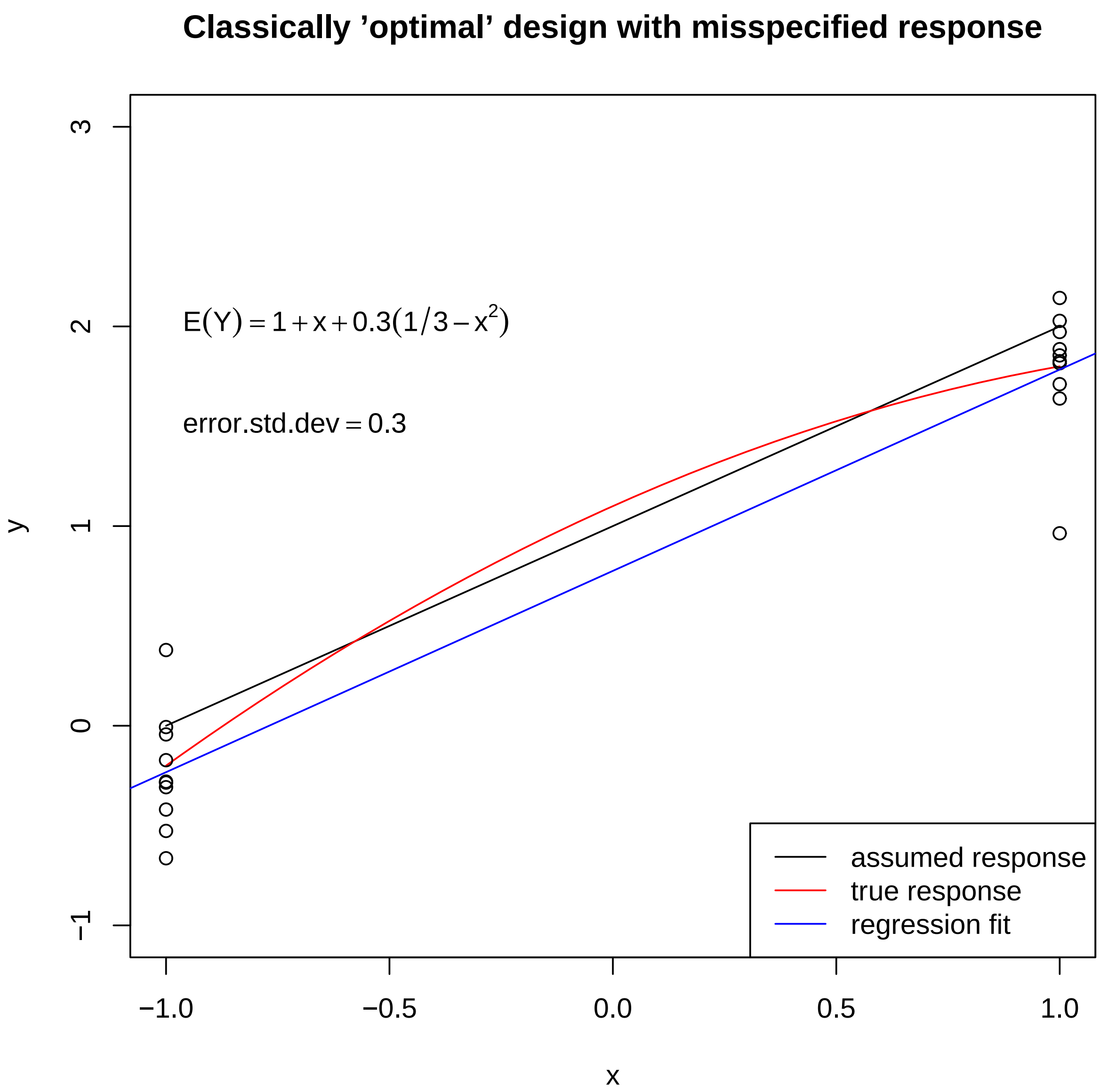
$$E\left[Y\left(x\right)\right]=\boldsymbol{f}'\left(x\right)\boldsymbol{\theta}+\psi\left(x\right),$$

for some 'small' function ψ.

  - Example: $\psi\left(x\right)=\alpha\cdot\left(1/3-x^2\right)$ for some 'small' α – quadratic 'contamination', orthogonal to $\boldsymbol{f}$:

$$\int_{-1}^{1}\boldsymbol{f}\left(x\right)\psi\left(x\right)dx=\begin{pmatrix}0\\0\end{pmatrix};\tag{1}$$

    this can always be arranged and ensures that $\boldsymbol{\theta}$ is well defined.



Classically 'optimal' design with misspecified response
$E(Y) = 1 + x + 0.3(1/3 - x^2)$
error.std.dev = 0.3
assumed response
true response
regression fit
y
x
−1.0
−0.5
0.0
0.5
1.0
−1
0
1
2
3

- Let ξ be the 'design', sometimes viewed as a pmf $\xi(x_i) = n_i/n$ if n_i observations are to be made at the design point x_i, at other times viewed as the 'design measure' or the edf of the design points. For instance the lse $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ has covariance matrix $\sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}$

$$
\begin{aligned}
&= \frac{\sigma_\varepsilon^2}{n}\left(\sum \frac{n_i}{n}\boldsymbol{f}(x_i)\boldsymbol{f}'(x_i)\right)^{-1} = \frac{\sigma_\varepsilon^2}{n}\left(\sum \xi(x_i)\boldsymbol{f}(x_i)\boldsymbol{f}'(x_i)\right)^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\mathbf{M}_\xi^{-1}, \text{ where } \mathbf{M}_\xi = \int_{-1}^{1}\boldsymbol{f}(x)\boldsymbol{f}'(x)\xi(dx).
\end{aligned}
$$

- If $E[Y(x)] = \boldsymbol{f}'(x)\boldsymbol{\theta} + \psi(x)$ then the presence of ψ introduces a bias:

$$
E\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \mathbf{M}_\xi^{-1}\mathbf{b}_{\psi,\xi}, \text{ where } \mathbf{b}_{\psi,\xi} = \int_{-1}^{1}\boldsymbol{f}(x)\psi(x)\xi(dx).
$$

  - Compare (1) – the continuous uniform 'design' is unbiased. ['Space-filling']

  - The mse matrix of $\hat{\boldsymbol{\theta}}$ is

$$
\mathsf{MSE}(\psi,\xi) = \frac{\sigma_\varepsilon^2}{n}\mathbf{M}_\xi^{-1} + \mathbf{M}_\xi^{-1}\mathbf{b}_{\psi,\xi}\mathbf{b}'_{\psi,\xi}\mathbf{M}_\xi^{-1}.
$$

- For a function $\phi(\cdot)$ like the determinant ('D-optimality'), trace ('A-optimality'), max eigenvalue ('E-optimality') etc. we might aim to minimize (the worst case of) $\phi(\text{MSE}(\psi, \xi))$:

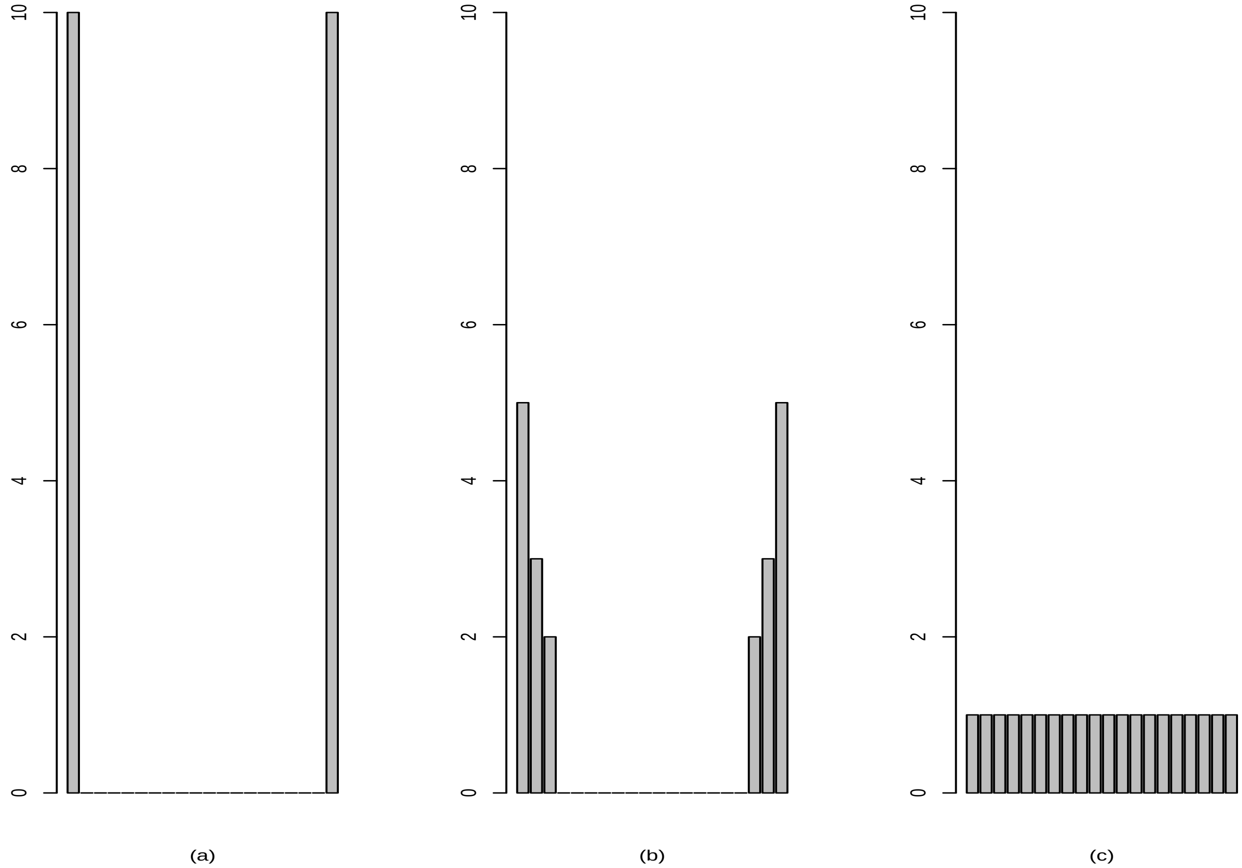
$$\min_{\xi} \max_{\psi} \phi(\text{MSE}(\psi, \xi)).$$

  – The max value $\phi(\text{MSE}(\psi_{\text{max}}, \xi))$ turns out to depend only on the maximum eigenvalue of this mse matrix (or one closely related to it); these eigenvalues are functions of the design itself.

- Back to SLR - if (as has always been assumed in these problems) the design is symmetric ($\xi(x) = \xi(-x)$) then

$$\mathbf{M}_\xi = \int_{-1}^{1} \boldsymbol{f}(x) \boldsymbol{f}'(x) \xi(dx) = \int_{-1}^{1} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \xi(dx) = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(\xi) \end{pmatrix},$$

and it turns out that the other matrices involved in MSE(ψ_{max}, ξ) are also diagonal, so that the eigenvalues can be read off; now proceed to minimize the larger (?) of them.



Implementations of minimax designs, $n = 20$; max subject to $\int \psi^2(x)dx \leq \tau^2/n$. (a) $\tau^2 = 0$, (b) τ^2 positive but small, (c) $\tau^2 \to \infty$.

- **Problem 1**: Does symmetry have to be assumed, or is the optimal ('minimax') design <u>necessarily</u> symmetric?
  - Possible route to a solution: Drop this assumption and determine the optimal design without it. This is complicated (for me) – the various matrices determining the loss, which are diagonal under symmetry, are no longer so.
  - Sometimes, as for D-optimality, there are nice convexity properties – the loss is smaller at the average $\left(\xi_+ + \xi_-\right)/2$ of the design for 'x' and the design for '$-x$', and this average design is symmetric. Not true for A- or E-optimality (although it is in 'classical' theory).
  - For the 'classical' theory all of these loss function are convex and symmetry is indeed necessary.

- **Problem 2**: For D-optimality, the optimal design is range-invariant: if the regression range $[-1, 1]$ is changed to $[a, b]$ then we linearly transform each design point x_i and put the same number of observations at the transformed design point as there were at x_i. This is not so for (classical or robust) A- and E-optimality – the design <u>weights</u> will also change. <u>How</u> do they vary with the range?

  – A solution to Problem 2 might shed light on the next problem – perhaps, in the case of SLR, it is a problem only for certain ranges $[a, b]$ and not for others.

- **Problem 3**: The minimax problem for SLR described above was reduced to finding an optimal design ξ_0 minimizing the maximum of two quantities:

$$\max\left(\lambda_1\left(\xi\right), \lambda_2\left(\xi\right)\right).$$

- What are these quantities? First (recall $\int \psi^2(x)dx \leq \tau^2/n$) define

$$\nu = \frac{\tau^2}{\tau^2 + \sigma_\varepsilon^2} \in [0, 1],$$

so $\nu = 0 \Rightarrow$ only variance of interest, $\nu = 1 \Rightarrow$ only bias of interest. Put $\mu_2 = \int x^2 \xi\left(dx\right)$. Let $m(x)$ be the density of the 'design' ξ (!) – the optimal design has to be approximated in the end, to implement it. Then – for instance – in the case of E-optimality and the regression range $\chi = [-1/2, 1/2]$ we have

$$\begin{aligned}
\lambda_1\left(\xi\right) &= 1 - \nu + \nu\left[\int_\chi \left(m\left(x\right) - 1\right)^2 dx\right], \\
\lambda_2\left(\xi\right) &= \frac{1-\nu}{\mu_2} + \nu\left[\int_\chi x^2\left(\frac{m\left(x\right)}{\mu_2} - 12\right)^2 dx\right].
\end{aligned}$$

- We are to minimize $\max\left(\lambda_1(\xi), \lambda_2(\xi)\right)$. A relatively easy route to a solution – <u>when it works</u> – is:

1. Hope that λ_1 will be the largest one, find ξ_1 minimizing $\lambda_1(\xi)$, verify that $\max\left(\lambda_1(\xi_1), \lambda_2(\xi_1)\right) = \lambda_1(\xi_1)$.

2. If not, do it with λ_2: find ξ_2 minimizing $\lambda_2(\xi)$, verify that $\max\left(\lambda_1(\xi_2), \lambda_2(\xi_2)\right) = \lambda_2(\xi_2)$.

BUT this doesn't always work; for A- and E-optimality, often

$$\begin{aligned}
\max\left(\lambda_1(\xi_1), \lambda_2(\xi_1)\right) &= \lambda_2(\xi_1) \text{ and} \\
\max\left(\lambda_1(\xi_2), \lambda_2(\xi_2)\right) &= \lambda_1(\xi_2).
\end{aligned}$$

- Shi, Ye, Zhou (2003) – non-smooth optimization methods (maximum eigenvalues aren't 'smooth') – obtain a description of the solution; severe computational problems.

- Alternate approach, being attempted for SLR:

1. Find ξ_1 minimizing $\lambda_1(\xi)$ in the class for which $\lambda_1(\xi) > \lambda_2(\xi)$.

2. Find ξ_2 minimizing $\lambda_2(\xi)$ in the class for which $\lambda_2(\xi) > \lambda_1(\xi)$.

Choose the one with the smaller maximum loss: put

$$\xi_0 = \begin{cases} \xi_1, & \text{if } \lambda_1(\xi_1) < \lambda_2(\xi_2), \\ \xi_2, & \text{otherwise.} \end{cases}$$

- Even if this works out for SLR it remains a problem for more complex models – approximate quadratic regression is the 'next' one.

- Problem 3 is not related to the non-invariance of A- and E-optimality under changes of the regression range. For quadratic regression it also arises in the case of D-optimality, even though this criterion <u>is</u> range-invariant.

- Sometimes the best 'solution' is to avoid the problem altogether:

  – Restrict the class of allowable designs - e.g. tractable parametric classes with the parameters optimally chosen.

  – Finite design space? – Avoids the problem of designs with densities (designs are now equivalent to vectors); integer optimization problems.

- A sampling of other problems in robustness of design:
  - Spatial sampling – look at robustness of the design against the ‘wrong’ spatial correlation function (as well as against the wrong response function); applications in computer experiments too.
  - Survey sampling - survey samplers resist model-based designs, but could perhaps be convinced otherwise if the models were sufficiently flexible.
  - Response surface exploration – sequentially search for a maximum on a surface, allowing for model errors.
  - Threshold designs – search for a level of the inputs at which the response exceeds a particular threshold (in the face of model uncertainty).

There is no shortage of problems to be ‘robustified’.