

Sequential M-MED

Kristyn Pantoja

10/04/2019

Beating quadratic D-Optimal design

B_n

$MSE(B_n)$

Posterior Probabilities of Hypotheses

Other Designs for Model Selection (T, Ds-Optimal, step-wise F-test)

Adjusting Sequential M-MED (unsuccessfully)

Larger alpha

some notes

Beating quadratic D-Optimal design

How, it seems, to beat quadratic D-Optimal Design

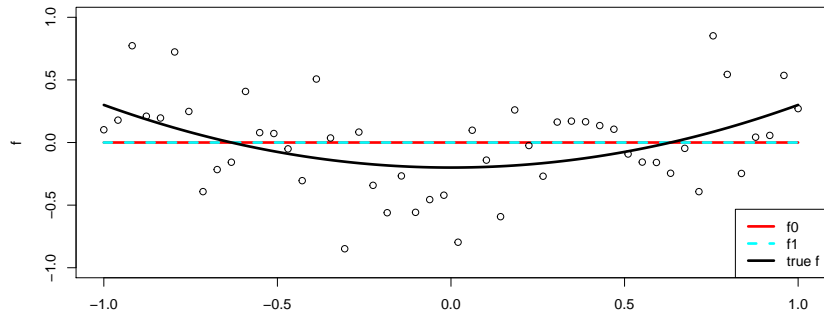
We want to come up with examples where the sequential M-MED gives higher prior probability of H_1 than the quadratic D-Optimal design does.

- ▶ It seems to me that when $\text{sigmasq} > \text{sigmasq01}$ (but not too big (??)), sequential M-MED typically does better than quadratic D-Optimal design in this sense.
- ▶ Is this sensible, though? To have a small value of sigmasq01 when the priors are wrong, anyway?
 - ▶ Is there a way to not allow the prior means to affect the calculation for posterior probabilities in nested models? non-informative / improper priors? Or is this not an issue when the prior means are 0 vectors - because then the marginal variance only matters in model evidence computation? (Otherwise, it seems it would be an issue because of how much it affects model evidence)
 - ▶ What does the ratio have to be?

Revisiting the Example

```
mu0 = c(0,0)
mu1 = c(0,0,0)
sigmasq01 = 0.025
sigmasq = 0.125 # 0.025 is too small
                # 0.075 is barely too small
                # 0.1 works, barely
                # 0.5 is too large (but BF01 is smaller)

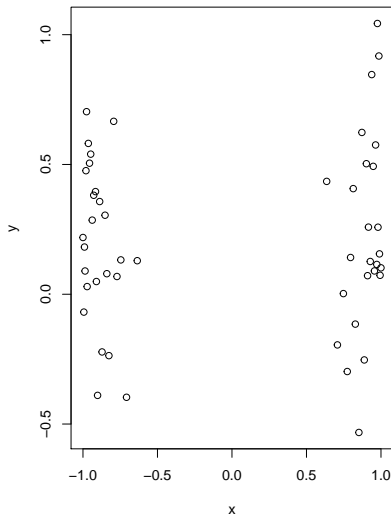
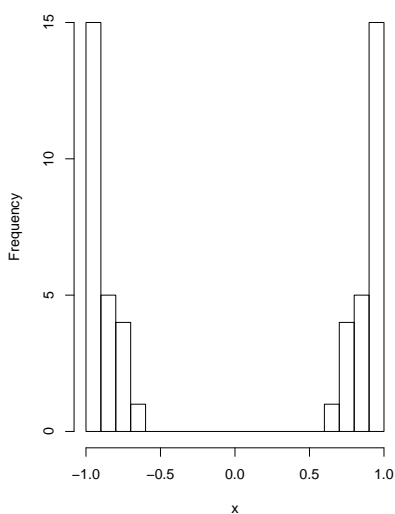
betaT = c(-0.2, 0, 0.5) # like ex2.1
```



M-MED Without Data

I generate 50 design points from M-MED with no data, one with $\alpha = 2p$.

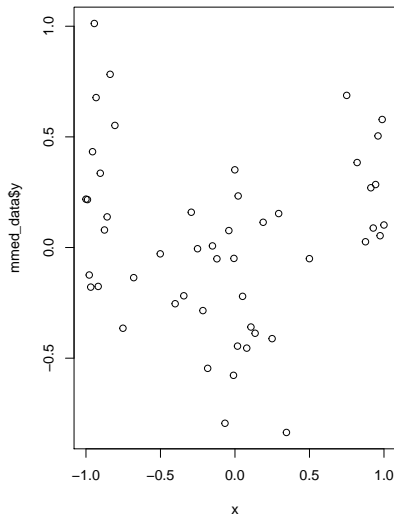
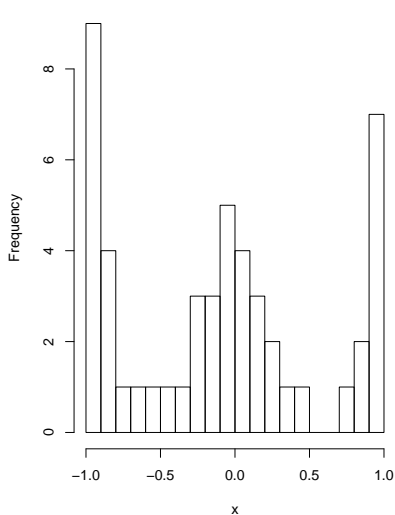
M-MED, alpha=2p



M-MED With Data

I work with a sequence of 5 steps, generating 10 points in each step, and resulting in 50 points, too.

M-MED, with data

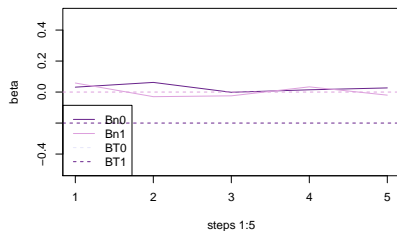


Bn

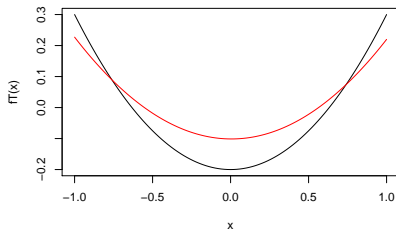
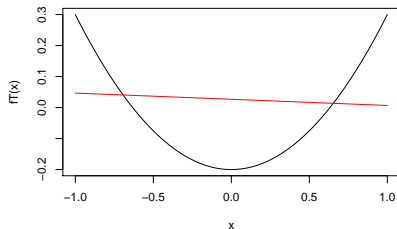
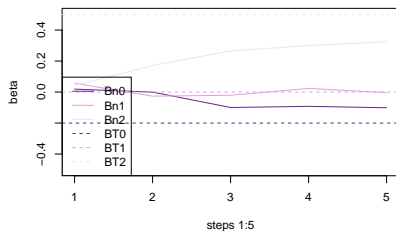
Seeing How Bn Changes

seems also sensitive when sigmasq01 is large (when sigmasq01 was 1, the quadratic term's coefficient was nearly 0, - and some other bad things happen I'm sure)

H0 Bn Estimates

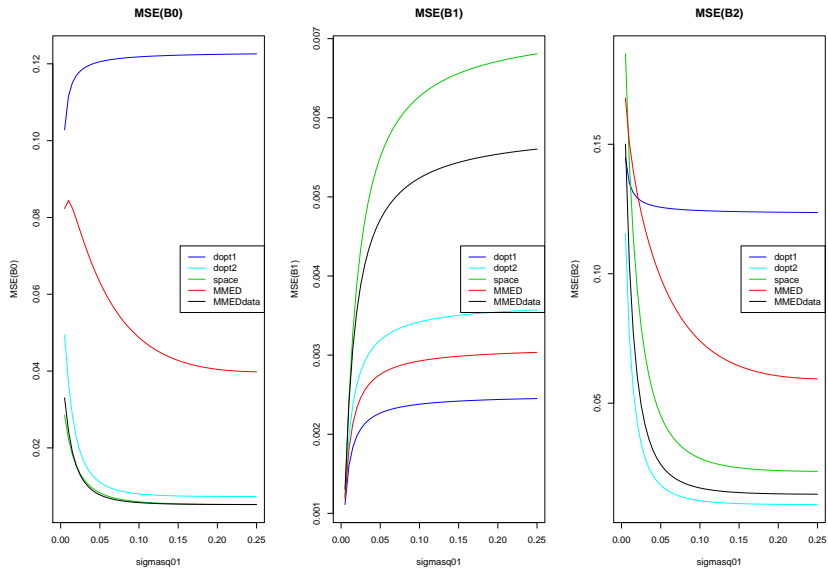


H1 Bn Estimates

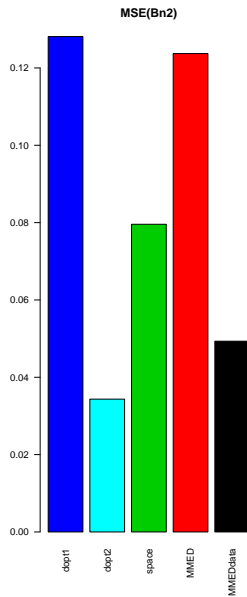
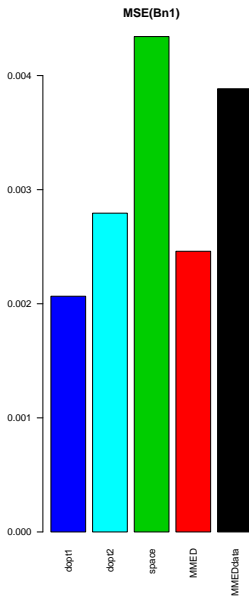
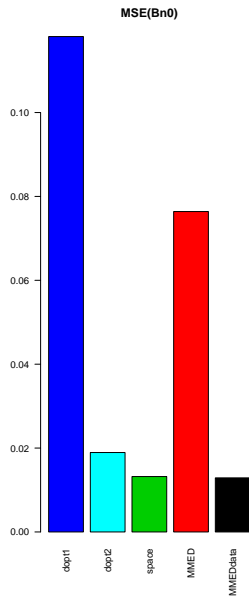


$$\text{MSE}(\mathbf{B}_n)$$

MSE(Bn) (quadratic model)

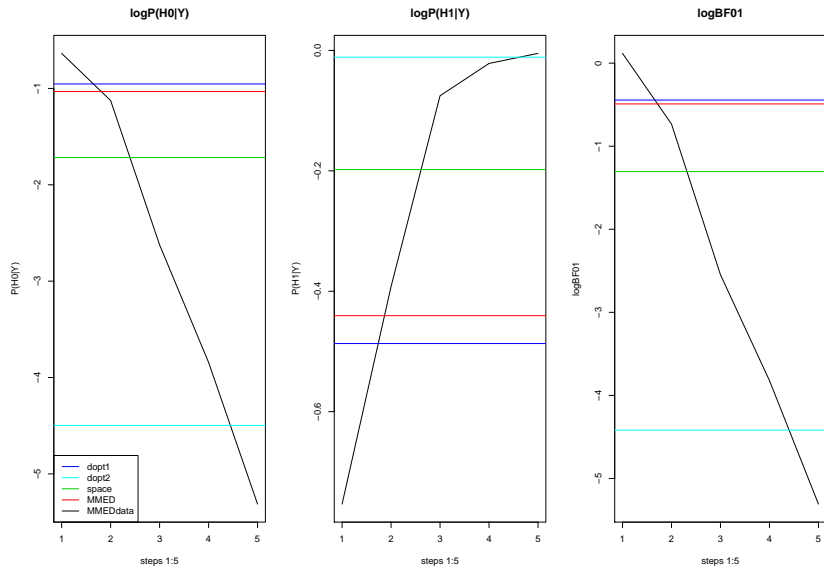


MSE Bar Plots



Posterior Probabilities of Hypotheses

Hypothesis Testing



	$E[P(H_0 Y) D]$	$E[P(H_1 Y) D]$	$E[BF_{01} Y,D]$
dopt1	0.38542	0.61458	0.64115
dopt2	0.01114	0.98886	0.01206
space	0.17955	0.82045	0.27095
MMED	0.35631	0.64369	0.61184
MMEDdata	0.00492	0.99508	0.00495

Other Designs for Model Selection (T,
Ds-Optimal, step-wise F-test)

T-Optimal Design

- ▶ Atkinson & Fedorov 1975 Biometrika “The Designs of Experiments for Discriminating between Two Rival Models”
- ▶ To compare models $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$ (not necessarily linear in the parameters),
 1. suppose that one model, $\eta_1(\cdot, \theta_1)$ for example is true
 2. choose design points x to maximize the sum of squares for lack of fit of the second model
- ▶ may be sensitive to choice of the fixed model and, depending on that, the choice of its parameters
 - ▶ if they are nested and the smaller model is true, then the design does not depend on any values of unknown parameters
 - ▶ otherwise, it does, because we must fix some parameter values for those parameters that are not included in the smaller model
- ▶ the focus is on model discrimination, not parameter estimation
- ▶ questions:
 - ▶ why are the design points, called support points, fixed (3 for quadratic models)? is this always the case?
- ▶ a problem: “Analytical methods of construction are possible only in a few special cases”

Ds-Optimal Design

- ▶ Stigler 1971 JASA “Optimal Experimental Design for Polynomial Regression”
- ▶ Used to discriminate between two nested polynomial models $y(x) = \sum_{i=1}^{m_0} \beta_i x^i$ and $y(x) = \sum_{i=1}^{m_1} \beta_i x^i$, $m_0 < m_1$
- ▶ “the model is thought to be of degree r , but possibly the coefficients of the $s = m_1 - m_0$ higher powers are not zero” (Studden, “Some Robust-Type D-Optimal Designs in Polynomial Regression”)
- ▶ Let

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where M_{11} is a dimension m_0 matrix, and M_{22} is a dimension s matrix.

- ▶ The problem is to find the design ξ that maximizes $|M_{11}(\xi)|$ subject to $|\Sigma_s(\xi)| \geq c$.
- ▶ the focus seems to be on parameter estimation, with some protection/robustness

Step-wise F-test

- ▶ Biswas & Chaudhuri 2002 Biometrika, “An efficient design for model discrimination and parameter estimation in linear models”
- ▶ A step-wise F-test to select the number of monomials that should be included.
- ▶ 3 objectives in this method:
 - ▶ selection of the correct (polynomial) model
 - ▶ efficient estimation of all parameters in that model
 - ▶ generation of design points so as to converge to optimal design for that model (optimal in what sense?)

Adjusting Sequential M-MED (unsuccessfully)

Why?

The D_s -Optimal design does not compare sharp hypotheses, like the current version of sequential M-MED. It compares nested polynomials without specifying priors on the coefficients. To make the comparison more fair (and, I hope, also to make sequential M-MED more robust and better-performing), we can adjust our current implementation to dynamically update the priors, too. I think this makes sense, since, as more data is gathered, we learn more about the true shape of the data.

Generating Sequential M-MED with data

Consider a sequence of T steps:

1. Choose N_1 initial points using M-MED (without data), with $\alpha = 0$; call it D_1 . This amounts to initializing with a space-filling design.
2. Collect data on D_1 ; call it y_1 .
3. Choose N_2 additional points using M-MED (with data), with $\alpha = \frac{2p}{T}t$, $t = 1 : T$; call these design points D_2 .
4. Collect data, y_2 , on D_2 .
5. Repeat steps 3-4 for $t = 2 : T$.

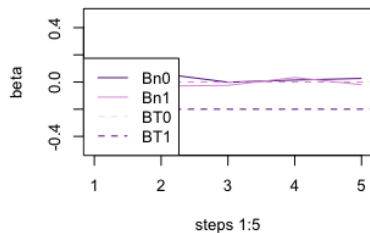
The part that would change is step 3 (and its iterations). Instead of using the posterior predictive distribution with the original hypothesized prior means $\mu^{(i)}$ for $i = \{0, 1\}$, we can let $\mu^{(i)} = \beta_n^{(i)}$ at each new iteration.

Issues

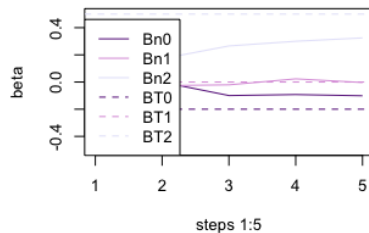
- ▶ This adaptive-prior version doesn't actually do better than the sequential M-MED, much less quadratic D-Optimal design :(
- ▶ However, in general, our method has the advantage of being easier to implement than other designs, since there is no need to maximize any determinants, which may require exchange algorithms

Bn estimates for Beta

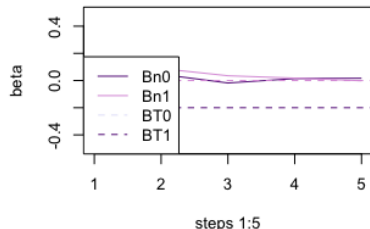
H0 Bn Estimates



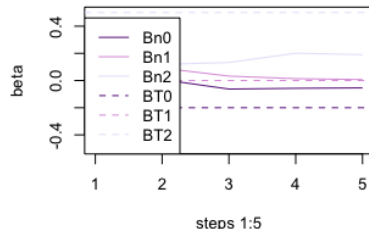
H1 Bn Estimates



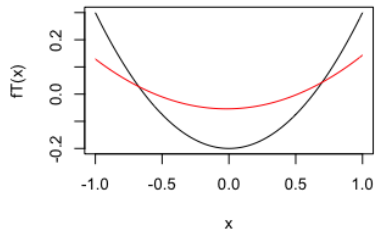
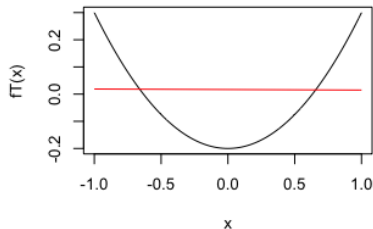
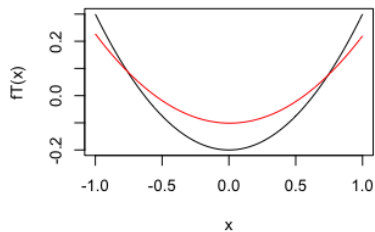
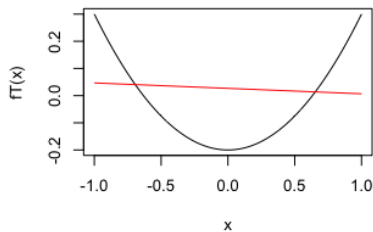
H0 Bn Estimates



H1 Bn Estimates

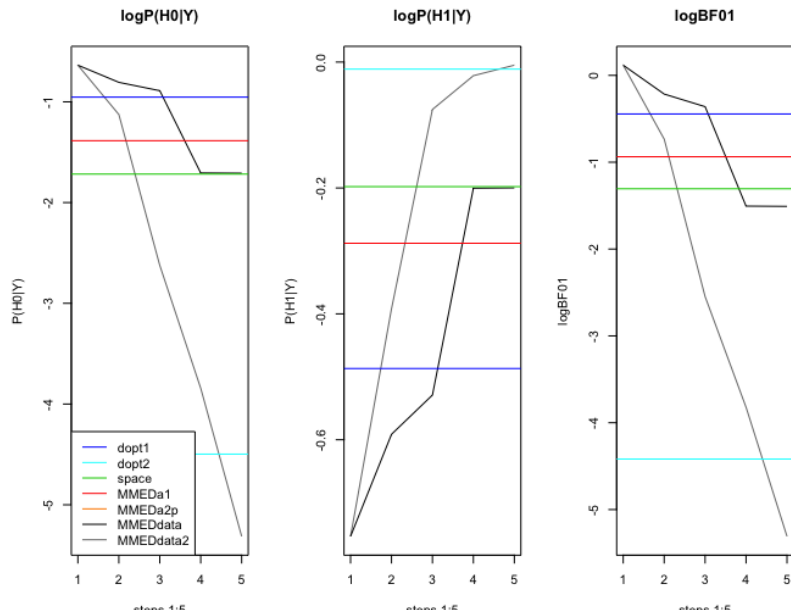


Bn for reconstructing f

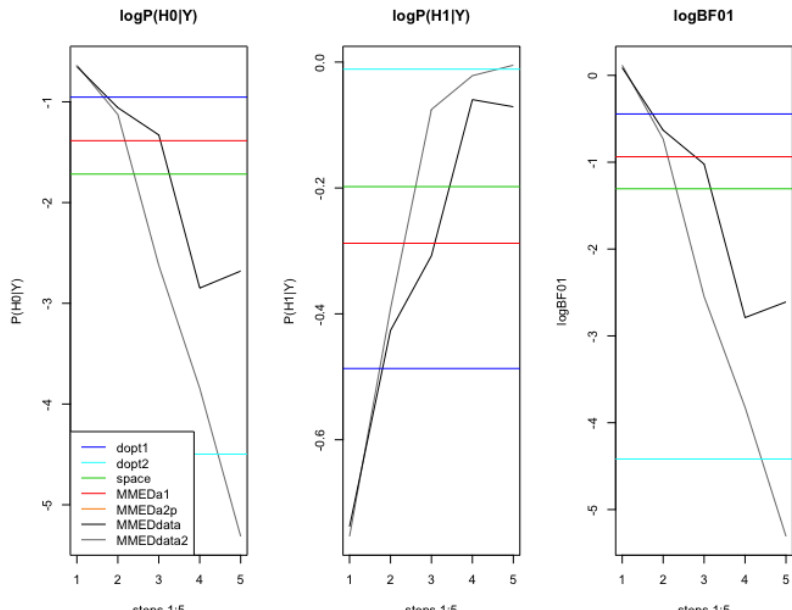


Note that the following pictures for posterior probabilities of H_1 are misleading because they assume different priors, or changing priors. In general, though, how can we correct this for non-sharp hypotheses for nested models? Bayesian methods already account for this, in fact. Maybe larger prior variance would make more sense, though.

Posterior Probabilities (original priors)



changing prior for adaptive sequential M-MED

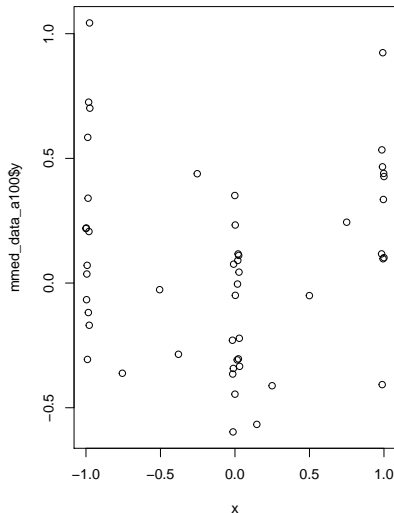
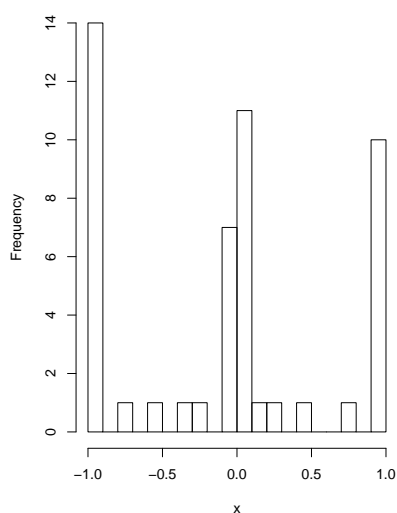


Larger alpha

$\alpha = 100$

very similar to quadratic D-Optimal design

M-MED, with data



some notes