

# Meeting Update

## Summaries

Kristyn Pantoja

Department of Statistics  
Texas AM University

18 April 2019

# Outline

## Last Time

## Sequentially Chosen Candidates in One-at-a-Time Algorithm

- Expected posterior probabilities of hypotheses

- Another metric, using distances in both  $x$  and  $y$

## Timing Things

## Comments

Last Time

# Last Time

## What happened last time

1. Started Gaussian Process Model Selection
2. Saw results for large  $K$

## Things to do:

1. Some statistics/numbers for evaluating the criterion, for the purpose of doing comparisons with other designs. Basically a way to show that the design is better!
2. Time things!
3. Clean up code, address pull request comments.

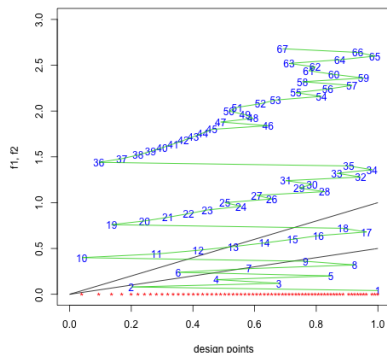
# Sequentially Chosen Candidates in One-at-a-Time Algorithm

# Uniformly vs. Sequentially Chosen Candidates for One-at-a-Time Algorithm

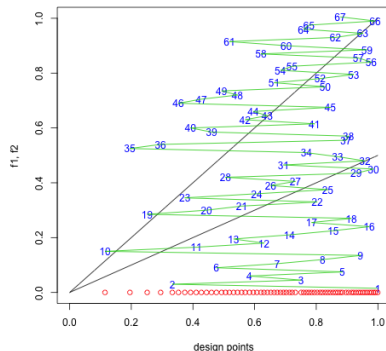
Here, *numCandidates* = 1500 in One-at-a-Time Algorithm, with candidates picked uniformly.

(Even when sorting uniformly-selected candidates)

Fast Algorithm K=100



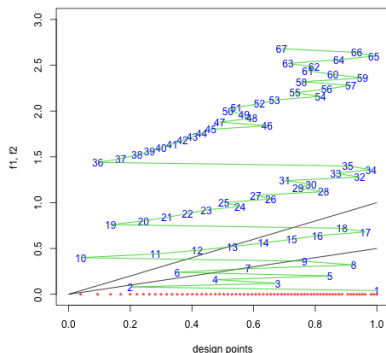
One-at-a-Time Uniform



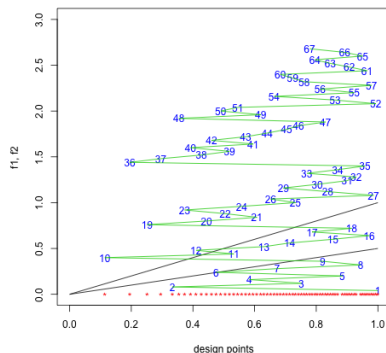
# Uniformly vs. Sequentially Chosen Candidates for One-at-a-Time Algorithm

When we choose candidates along a sequence instead of uniformly, the two designs are closer. (Why? Rui Tui's seminar talk, when he mentioned difference in sphere packing)

Fast Algorithm K=100



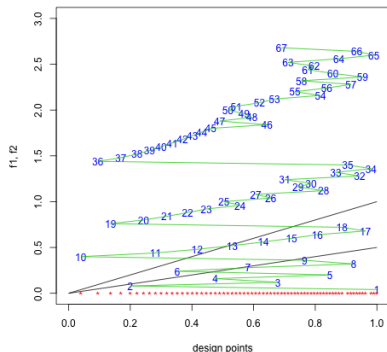
One-at-a-Time Sequential



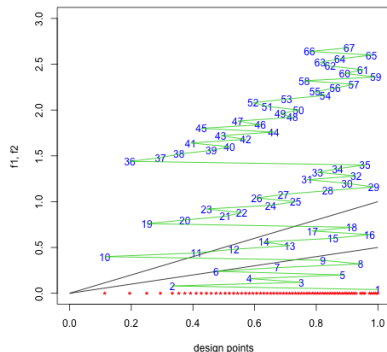
# Uniformly vs. Sequentially Chosen Candidates for One-at-a-Time Algorithm

Here,  $\text{numCandidates} = 10^5$  in One-at-a-Time Algorithm. They're even closer!

Fast Algorithm K=100



One-at-a-Time Sequential

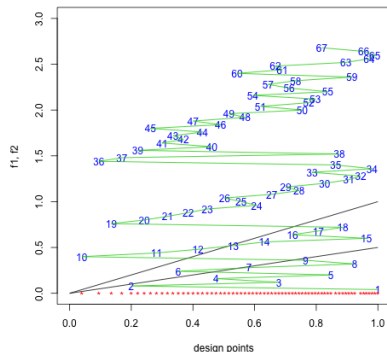




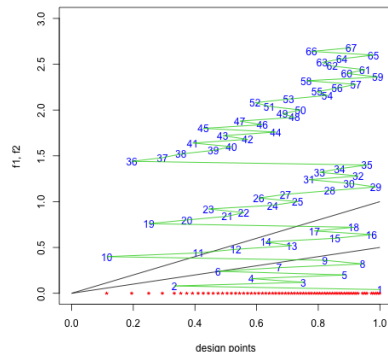
# Uniformly vs. Sequentially Chosen Candidates for One-at-a-Time Algorithm

Compare Fast with  $K = 20$  and One-at-a-Time with  $numCandidates = 10^5$  in One-at-a-Time Algorithm.

## Fast Algorithm K=20



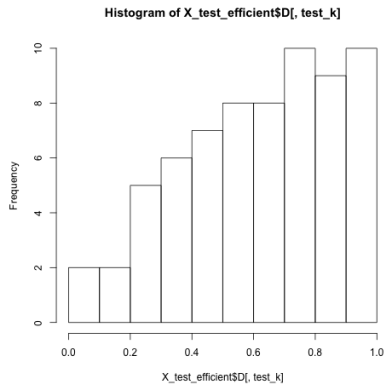
## One-at-a-Time Sequential



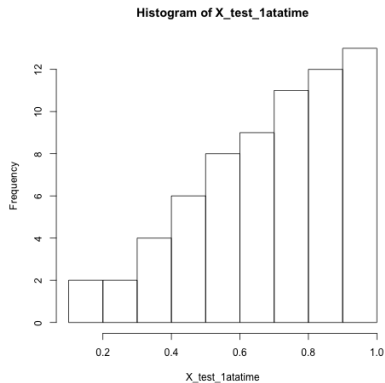
# Histograms for Designs

Comparing designs in Fast ( $K = 100$ ) and One-at-a-Time with candidates along sequence ( $numCandidates = 10^5$ )

## Fast



## One-at-a-Time Seq



# Summary Stats on Designs

Comparing designs in Fast ( $K = 100$ ) and One-at-a-Time with candidates along sequence ( $numCandidates = 10^5$ )

## Fast

1. mean, sd of design points = 0.6129598, 0.2540533
2. sum, mean, sd of differences between consecutive sorted points = 0.9599863, 0.01454525, 0.00766221
3. sum, mean, sd of differences between  $f_0$  and  $f_1$  = 20.53415, 0.3064799, 0.1270267

## One-at-a-Time

1. mean, sd of design points = 0.6882634, 0.2182556
2. sum, mean, sd of differences between consecutive sorted points = 0.8859489, 0.01342347, 0.01187598
3. sum, mean, sd of differences between  $f_0$  and  $f_1$  = 23.05683, 0.3441317, 0.1091278

# Thoughts

Comparing designs in Fast ( $K = 100$ ) and One-at-a-Time with candidates along sequence ( $numCandidates = 10^5$ ), it seems....

1. The fast algorithm is more space-filling
2. Its mean is smaller... not sure what this implies. Not as focused on distinguishing the models? More for prediction/parameter estimation?
3. ...

# Sequentially Chosen Candidates in One-at-a-Time Algorithm

Expected posterior probabilities of hypotheses

## Expected Posterior Probability of Model

- ▶ Posterior probability of a model given by  $H_\ell, \ell = 1, \dots, M$  to see which model is more likely to be the correct one.
- ▶ Posterior Probability is

$$P(H_\ell|Y) = \frac{\pi_\ell P(Y|H_\ell)}{\sum_{m=1}^M \pi_m P(Y|H_m)} \quad (1)$$

where  $\pi_m$  is the prior on the model given by  $H_m$ , and  $P(Y|H_m)$  is the model evidence.

- ▶ Consider the linear model,

$$y = f(\mathbf{x}) + \epsilon$$

$$f(x) = \mathbf{x}\beta$$

where  $\epsilon \sim N(0, \sigma_\epsilon^2)$  and  $\beta \sim N(\tilde{\beta}, \sigma_\beta^2)$ .

- ▶ For model  $H_\ell$  with  $\epsilon_\ell \sim N(0, \sigma_{\epsilon_\ell}^2)$  and  $\beta_\ell \sim N(\tilde{\beta}_\ell, \sigma_{\beta_\ell}^2)$ , we have  $y|H_\ell \sim N(\tilde{\beta}_\ell \mathbf{x}, \sigma_{\epsilon_\ell}^2 + \mathbf{x}^2 \sigma_{\beta_\ell}^2)$  where the mean and variance are computed by iterated expectation and variance.
- ▶ Since we don't have  $Y$  to calculate the model evidence, instead calculate the expected model evidence  $E_Y[P(Y|H_m)]$

## Estimate Expected Posterior Probability of Model

1. Obtain design  $\mathbf{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  from the Fast Algorithm for Linear Model Selection.
2. Draw  $y_i^{(j)} \sim N(\tilde{\beta}_\ell \mathbf{x}_i, \sigma_{\epsilon_\ell}^2 + \mathbf{x}_i^2 \sigma_{\beta_\ell}^2)$ ,  $\forall \mathbf{x}_i \in \mathbf{D}, j = 1, \dots, J$  to obtain  $J$  simulations of  $Y = \{y_1, \dots, y_N\}$  for model  $H_\ell$ , which gives the linear model:

$$y = f(\mathbf{x}) + \epsilon$$

$$f(\mathbf{x}) = \mathbf{x} \beta_\ell$$

where  $\epsilon \sim N(0, \sigma_{\epsilon_\ell}^2)$  and  $\beta_\ell \sim N(\tilde{\beta}_\ell, \sigma_{\beta_\ell}^2)$ .

3. Estimate  $E_Y[P(Y|H_m)] \approx \frac{1}{J} \sum_{j=1}^J P(Y|H_m)$ ,  $\forall m \in \{1, \dots, M\}$ , where  $P(Y|H_m)$  is the pdf of the distribution  $N(\tilde{\beta}_m \mathbf{x}, \sigma_{\epsilon_m}^2 + \mathbf{x}^2 \sigma_{\beta_m}^2)$ .
4. Then estimate  $E_Y[P(H_\ell|Y)]$  by

$$E_Y[P(H_\ell|Y)] = \frac{\pi_\ell E_Y[P(Y|H_\ell)]}{\sum_{m=1}^M \pi_m E_Y[P(Y|H_m)]} \quad (2)$$

## Questions about Step 3

- ▶ How to get a single number out of this computation?
- ▶ Where to put the design points  $\mathbf{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ?
- ▶ What do we do we do with  $y_i^{(j)}$  from simulations?
- ▶ Is there some kind of likelihood for linear regression? Some sources talk about Laplace approximation to marginal  $Y$ ... did I compute that wrong? I averaged over  $\beta$ , am I supposed to somehow average over  $\mathbf{D}$  as well?
- ▶ Need to get rid of  $\mathbf{x}$ 's somehow...



# What I Ended Up Doing

1. Obtain design  $\mathbf{D}$  from the Fast Algorithm for Linear Model Selection.
2. Draw  $y_i^{(j)} \sim N(\tilde{\beta}_\ell \mathbf{x}_i, \sigma_{\epsilon_\ell}^2 + \mathbf{x}_i^2 \sigma_{\beta_\ell}^2), \forall \mathbf{x}_i \in \mathbf{D}, j = 1, \dots, J$  for model  $H_\ell$ . For now,  $M = 2$ .
3. Estimate  $E_Y[P(Y|H_m)] \approx \frac{1}{JN} \sum_{j=1}^J \sum_{i=1}^N P(y_i|H_m, \mathbf{x}_i), \forall m \in \{1, \dots, M\}$ , where  $P(y_i|H_m, \mathbf{x}_i)$  is the pdf of  $N(\tilde{\beta}_m \mathbf{x}_i, \sigma_{\epsilon_m}^2 + \mathbf{x}_i^2 \sigma_{\beta_m}^2)$ .
4. Assume  $\pi_m = \frac{1}{M}, \forall m = 1, \dots, M$  and estimate  $E_Y[P(H_\ell|Y)]$  by

$$E_Y[P(H_\ell|Y)] \approx \frac{E_Y[P(Y|H_\ell)]}{\sum_{m=1}^M \pi_m E_Y[P(Y|H_m)]} \quad (3)$$

5. Since  $M = 2$ , can also compute the Bayes Factor,

$$BF_{01} \approx \frac{E_Y[P(Y|H_0)]}{E_Y[P(Y|H_1)]} \quad (4)$$

# New Results

For both  $H_0, H_1$ ,  $\sigma_\epsilon = 0.01, \sigma_\beta = 0.001, N = 51$ .  
 $\tilde{\beta}_0 = 1, \tilde{\beta}_0 = 1/2$ .

For generating  $Y$  under the  
null model given by  $H_0$ ,

- ▶  $E_Y[P(H_0|Y)] \approx 0.8196135$
- ▶  $E_Y[P(H_1|Y)] \approx 0.1803865$
- ▶  $BF_{01} \approx 4.543653$ : supports  $H_0$

For generating  $Y$  under the  
alternative model given by  $H_1$ ,

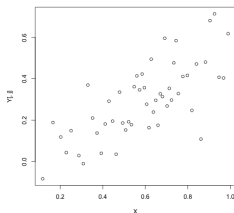
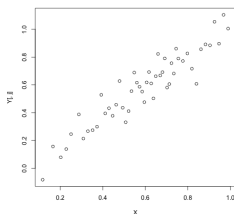
- ▶  $E_Y[P(H_0|Y)] \approx 0.1818883$
- ▶  $E_Y[P(H_1|Y)] \approx 0.8181117$
- ▶  $BF_{01} \approx 0.222327$ : supports  $H_1$

## Results, continued...

For both  $H_0, H_1$ ,  $\sigma_\epsilon = 0.01, \sigma_\beta = 0.001, N = 51$ .

$\tilde{\beta}_0 = 1, \tilde{\beta}_0 = 1/2$ .

Example Simulation from  $H_0$   
Model,



Interpreting Bayes Factors  
(Kass & Raftery 1995):

$BF_{01}$	Evidence for $H_0$
$< 1$	Supports $H_1$
1 to 3	not worth mentioning
3 to 20	positive
20 to 150	Strong
$> 150$	v strong

# Sequentially Chosen Candidates in One-at-a-Time Algorithm

Another metric, using distances in both  $x$  and  $y$

# Summary Stats on Designs

How do we use these to evaluate the design?

## Fast

1. mean, sd of design points = 0.6129598, 0.2540533
2. sum, mean, sd of differences between consecutive sorted points = 0.9599863, 0.01454525, 0.00766221
3. sum, mean, sd of differences between  $f_0$  and  $f_1$  = 20.53415, 0.3064799, 0.1270267

## One-at-a-Time

1. mean, sd of design points = 0.6882634, 0.2182556
2. sum, mean, sd of differences between consecutive sorted points = 0.8859489, 0.01342347, 0.01187598
3. sum, mean, sd of differences between  $f_0$  and  $f_1$  = 23.05683, 0.3441317, 0.1091278

# Timing Things

# Times

1. Fast Algorithm with  $K = 20$ : 2.77915 minutes
2. Fast Algorithm with  $K = 100$ : 58.4566 minutes
3. One-at-a-Time Sequential with  $numCandidates = 10^5$ : 17.37015 minutes
4. The "Fast" Algorithm is probably slower since  $K$  is so large... but it has to be.

## Comments



## Some Things.

1. Candidate set:  $N/2$  from space filling design in  $L_{jk}$ , and other  $N/2$  some linear combination of nearby points (Nesterov?) - get a total of  $N$  additional candidates in each new design  $k = 1, \dots, K$
2. So I suppose that would be the next step? Will hopefully see faster convergence.
3. Maybe choose a different power than  $2p = 2$  for  $d(.,.)$  to see if it makes the designs even more similar?
4.  $\gamma_{k+1} = \frac{k}{K-1}$  since we're indexing from  $k = 1, \dots, K-1$  to create the next design  $k = 2, \dots, K$ , resp.
5. Still thinking about asymptotic properties / limiting distribution (if it exists) for the fact that we're using Wasserstein distance instead of a density now.