# 1D Experiments (tuning), GP Ideas

Kristyn Pantoja

8/23/2019

Tuning q and f

Gaussian Process Model Selection

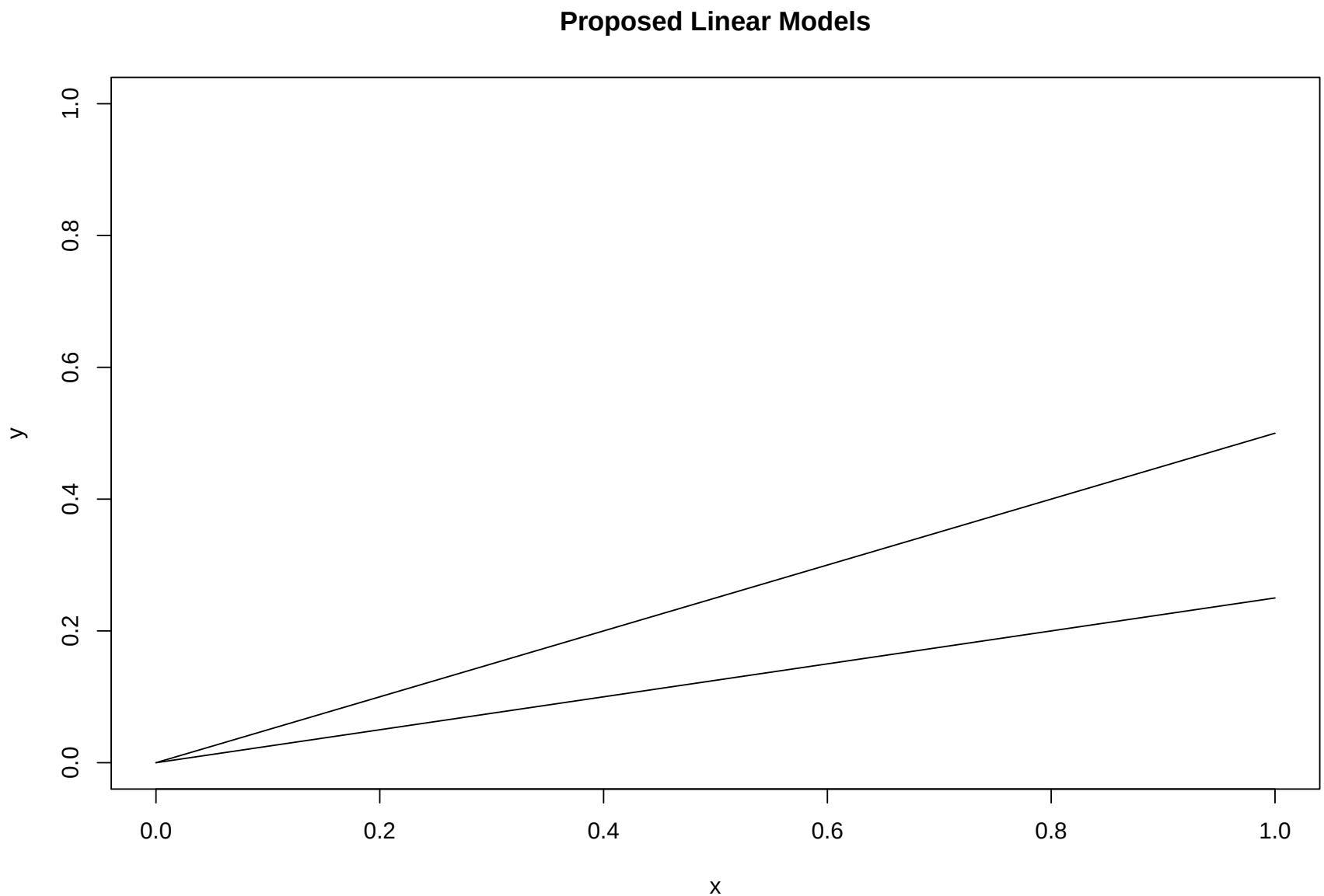
Adding Points to a Design using MED Methods

# Tuning q and f

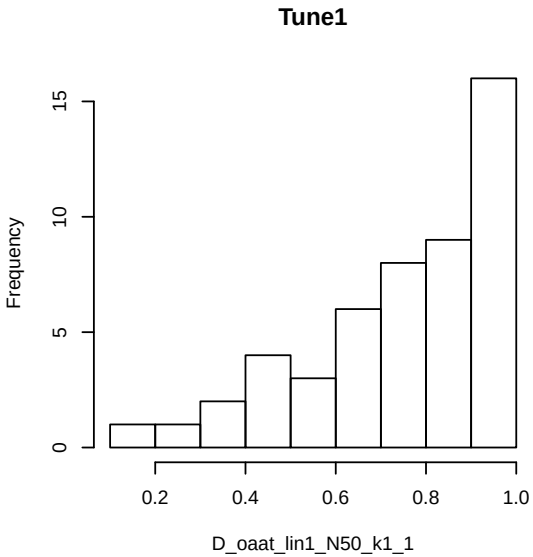
# Why tune

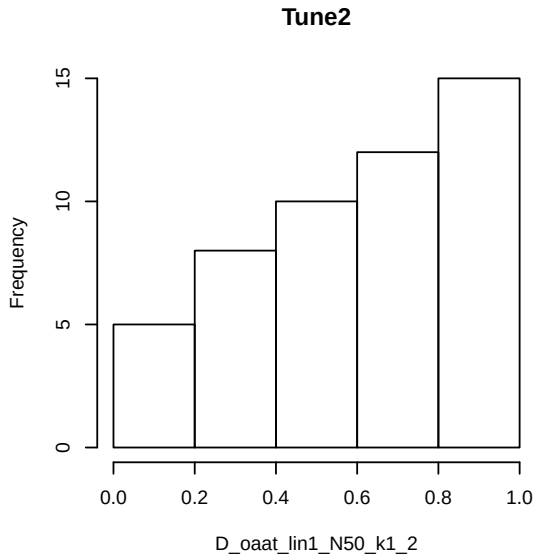
- To see if we can de-emphasize the space-filling tendency
- To see what happens when we use different definitions for f and q
- Using One-at-a-Time Algorithm for illustration, with 'k = 1' and 'N = 50'
- Designs where:
  1. $q = 1/W$
  2. $q = 1/W^{1/2p}$
  3. $q = 1/(W + c)^{1/2p}$, with $c = 0.5$
  4. $\mu_0 = \mu_0 + c$, with $c = 0.5$
  5. $\mu_\ell = \mu_\ell + c, \ell = \{0, 1\}$, with $c = 0.5$

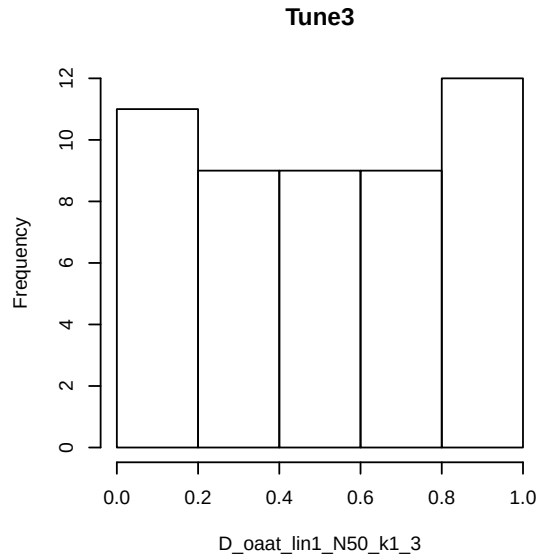
# I. Proposed Linear Models

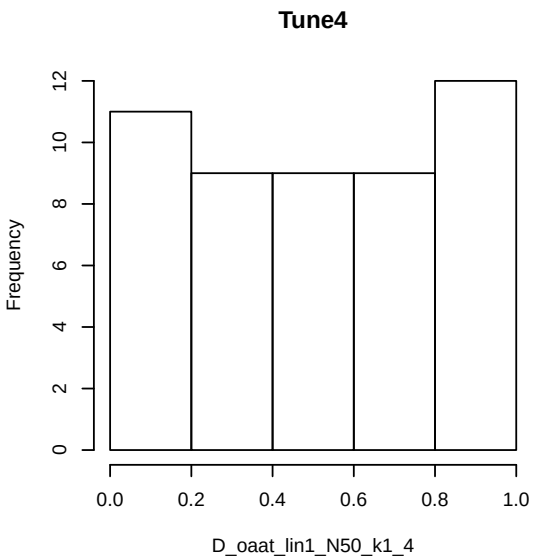


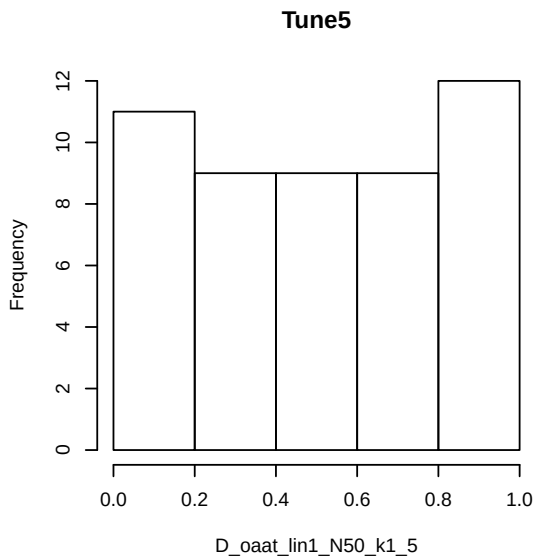
# Tuning Results



# Some Observations

- Getting rid of the $1/2p$ power on $f = W$ helps de-emphasize the space-filling aspect of the design
- Issues: What is the limiting distribution of the design? Even if we have the $1/2p$ power, what is the limiting distribution when $f$ is no longer a distribution?
    1. (Idea 1) Show that this method is asymptotically equivalent to that of Box and Hill, which "sequentially chooses points where the predicted responses are furthest apart": $\{\hat{y}_{1n}(x) - \hat{y}_{2n}(x)\}^2$, resulting in the oconstruction of "T-optimum designs that maximizes the noncentrality parameter in the sum of squares for testing lack of fit of the false model." (Atkinson 1981, "A Comparison of Two Criteria for the Design of Experiments for Discriminating Between Models")
    2. (Idea 2) Relate to sequential Bayesian likelihood ratio tests using bounds from KL divergence like in Chernoff's "Sequential Design of Experiments"

# Gaussian Process Model Selection

# Issues

- When a candidate point $\mathbf{x}_*$ are too close to the design points, correlation is too high, which leads to standard deviation for posterior predictive distribution $y_*|\mathbf{x}_*, X, Y$ to be too close to zero (computational issues).
  - The sign: fewer negative standard deviations when parameter $l$ is smaller
- After we choose one new design point, how do we update the posterior predictive distribution when choosing another design point? Interpolate at the newly chosen point? Use posterior predictive mean (but which one: $H_0$ or $H_1$)?

# One-at-a-Time Algorithm (2015)

Steps to obtain MED using One-at-a-Time algorithm:

1. Obtain *numCandidates* candidate points, $\mathbf{x}$, in $[0, 1]$ to form candidate set $C$.
2. Initialize $D_N$ by choosing $\mathbf{x}_1$ to be the candidate $\mathbf{x}$ which optimizes $f$, where $f(\mathbf{x}) = \text{Wasserstein}(\phi_{0,\mathbf{x}}, \phi_{1,\mathbf{x}})$ and

$$\phi_{0,\mathbf{x}} = N(\mu_0 \mathbf{x}, \sigma_0^2 + \mathbf{x}^2 \nu_0^2),$$
$$\phi_{1,\mathbf{x}} = N(\mu_1 \mathbf{x}, \sigma_1^2 + \mathbf{x}^2 \nu_1^2)$$

3. Choose the next point $\mathbf{x}_{j+1}$ by:

$$\mathbf{x}_{j+1} = \arg\min_{\mathbf{x} \in C} \sum_{i=1}^{j} \left( \frac{q(\mathbf{x}_i) q(\mathbf{x})}{d(\mathbf{x}_i, \mathbf{x})} \right)^k$$

where $q = 1/f^{(1/2p)}$, $d(x, y)$ is Euclidean distance and $k = 4p$.

# One-at-a-Time Algorithm for GP?

Suppose you have training data $X, Y$.

1. Obtain candidate set $C$
2. Initialize $\mathbf{D}$ as the candidate point $\mathbf{x}_*$ that maximizes $f(\mathbf{x}) = \text{Wasserstein}(\phi_{0,\mathbf{x}}, \phi_{1,\mathbf{x}})$, where, here, $\phi_{\ell,\mathbf{x}}$ is the predictive distribution $f_*|\mathbf{x}_*, X, f \sim N(k_*^T K^{-1} Y, k(\mathbf{x},\mathbf{x}) - k_*^T K^{-1} k_*)$, where $k_* = k(\mathbf{x}, X), K = K(X, X)$, and $k$ and $K$ are determined by the hypothesis $\ell$.
3. For subsequent design points, choose:

$$\mathbf{x}_{j+1} = \arg\min_{\mathbf{x}\in C} \sum_{i=1}^{j} \left( \frac{q(\mathbf{x}_i)q(\mathbf{x})}{d(\mathbf{x}_i,\mathbf{x})} \right)^k + \sum_{x_i \in X} \left( \frac{q(\mathbf{x}_i)q(\mathbf{x})}{d(\mathbf{x}_i,\mathbf{x})} \right)^k$$

where the data for previously added design points $\{(\mathbf{x}_i)|i = 1 : j\}$ are ...

# Adding Points to a Design using MED Methods

# What do we do about initializing the first point, considering the design has already been initialized?

- Check if the optimal point (that which maximizes the Wasserstein distance between $f_0$ and $f_1$) is in the set and proceed accordingly
- Don't initialize with the optimal point: skip straight to choosing the next design point based on the algorithm's criterion to approach minimizing total potential energy.