Stat 890
Design of computer experiments

- Last day: Introduced designs for computer experiments
- Saw that many common designs are motivated by Monte Carlo integration rather than computer model emulation
- Also saw that combining criteria seems to be helpful
- Good review paper Pronzato and Muller (2002)
- Today: more on design

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Uniform designs

- Another way to view space-filling is to want designs that spread points out uniformly in the input region, $[0,1]^d$
- Intuitively, would like the distribution of points to resemble a sample from a d-dimensional uniform distribution
- **Idea:** the design that has the minimum discrepancy between the empirical distribution function and the uniform cumulative distribution function
- Can use Kolmogorov Smirnov *discrepancy*

$$D(X) = sup_{x \in [0,1]} \left| F_n(x) - F(x) \right|$$

- or,

$$D_p(X) = sup_{x \in [0,1]} \left| \int |F_n(x) - F(x)|^p dx \right|^{1/p}$$

Comments

- Large literature for uniform designs usually aimed at numerical integration - quasi-Monte Carlo- (see Lemieux, 2009, text)
- Uniform designs are often very hard to find
- Fang et al. (2000) point out that uniform designs often have columns that are orthogonal or have small correlation
- So, can restrict search by combining criteria
- Can also look for uniformity within the class of Latin hypercube designs

Comments

- Another important argument for space-filling in general comes from the linear model

$$Y(X) = X\beta + \epsilon$$

- Suppose that there is only 1 factor and you want to estimate the mean and linear effect

$$Y(x) = \beta_0 + \beta_1 x + \epsilon$$

- Do you know what the optimal design is on [0,1]?

Comments

- Box and Draper (1959) pointed out that for polynomial regression with a misspecified degree, spreading out the design points can help minimize the bias with potentially little cost in prediction variance
- Wiens (1991) shows that uniform designs can be robust for certain departures from the regression model (*F*-tests)
- This is all to say, that space filling and uniformity are useful general properties

Model based criteria

- Suppose computer emulation is the desired goal
- Why not choose a design specifically designed to do this?
- What does it mean to emulate well?

Model based criteria

- Consider the GP framework
- The mean square prediction error is $MSPE(\hat{Y}(x)) = E\left[(\hat{Y}(x) - Y(x))^2\right]$
- Would like this to be small for every point in $[0,1]^d$
- So, criterion (Sacks, Schiller and Welch, 1992) becomes

$$IMSPE = \int_{x\in[0,1]^d} \frac{MSPE(\hat{Y}(x))}{\sigma^2} dx$$

- The optimal design minimized the *IMSPE*

Model based criteria

- Problems:
  - Integral is hard to evaluate
  - Function is hard to optimize
  - Need to know the correlation parameters
    - Could guess
    - Could do a 2-stage procedure aimed at guessing correlation parameters
    - Could use a Bayesian approach $BIMSPE = \int IMSPE \, dF(\theta)$

Model based criteria

- **Alternative:** Minimize the maximum mean square prediction error

$$\min_{X \in [0,1]^d} \max_{x \in X} MSPE(\hat{Y}(x))$$

- Again, need to know the correlation parameters to compute the criterion
- **Propose an algorithm to find a good design…**
- **Suppose you could do sequences of runs (batches), how would you do this algorithmically?**

Model based criteria

- There are more criteria than one could reasonably go through
- Other criteria include D-optimality and maximum entropy designs
- Interesting that the maximum entropy design (a measure of unpredictability of a random variable) criterion calls to maximize $\det(\sigma^2 R)$ for a design (Shewry and Wynn, 1987; Currin et al., 1991)

Sample size

- So you want to run a computer experiment… why?
- For numerical integration, there are often bounds associated with the estimate of the mean (e.g., Koksma-Hlawka theorem … see Lemieux, 2009) derives an upper bound on the absolute integration error
- For computer model emulation Loeppky, Sacks and Welch (2009) proposed a general rule for n=10d … well sort of

Sample size

- Two main considerations make this rule plausible:
  - The sensitivity of the model to inputs
  - Effect sparsity
- Several authors (Chapman et al., 1994; Jones et al., 1998) have suggested that *n*=10d has been a successful choice for the sample size in their experience
- Statistical model is the GP we have been looking at so far with constant mean

$$R(\mathbf{x}, \mathbf{x}') = \exp(-h(\mathbf{x}, \mathbf{x}')),$$

$$h(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \theta_j |x_j - x_j'|^{p_j},$$

$$\hat{Y}(\mathbf{x}) = E(Y(\mathbf{x})|\mathbf{y}, \boldsymbol{\theta}) = \hat{\mu} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})$$

Sample size

- Obviously would $\mathrm{MSE}(\hat{Y}(\mathbf{x}))$

$$= E(\hat{Y}(\mathbf{x}) - Y(\mathbf{x}))^2$$

$$= \sigma^2\left(1 - \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) + \frac{(1 - \mathbf{1}^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}))^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}\right)$$

- Note: $E\left|\frac{\partial Y(\mathbf{x})}{\partial x_j}\right|^2 = 2\sigma^2\theta_j$

- Interpretation?

Sample size

- Letting h be the weighted distance between two points in a random Latin hypercube design of size n

$$E(h) = m_1(n)\sum_{j=1}^{d}\theta_j \qquad \text{Var}(h) = m_2(n)\sum_{j=1}^{d}\theta_j^2$$

$$R(\mathbf{x}, \mathbf{x}') = \exp(-h(\mathbf{x}, \mathbf{x}')),$$

$$h(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d}\theta_j|x_j - x_j'|^{p_j},$$

- The find that MSE is impacted by the sum in both terms
  - MSE gets big if first term gets big, for example

- Consider the efficiency:

Sample size
Simulation

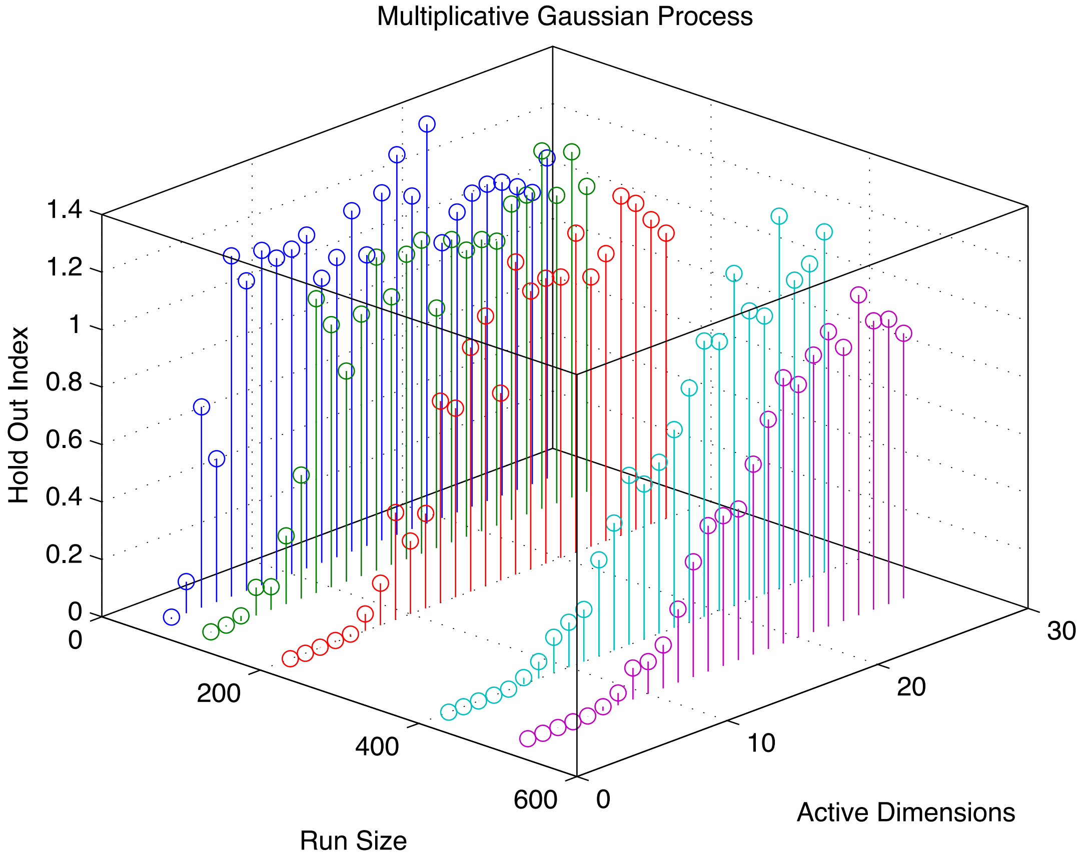
- Suppose you have a realization of a GP in d-dimensions and also a hold-out set for validation

- Consider the impact of more active dimensions

- Efficiency index:

$$HOI = \frac{MSPE_{ho}}{Var(Y_{ho})}$$

Sample size Simulation

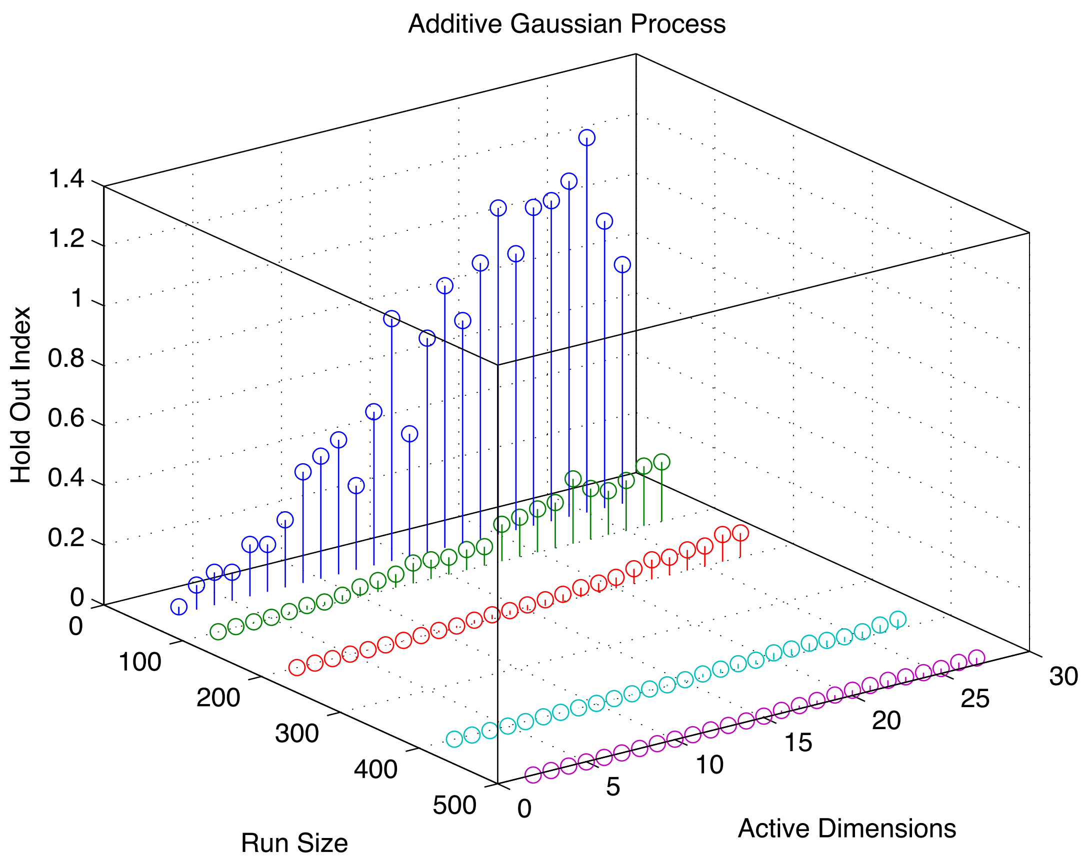


SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Sample size
Simulation

- Let's discuss what we see…

Sample size Simulation



SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Sample size
Simulation

- Let's discuss what we see…

Sample size
Simulation

- Paper by Loeppky et al. (2009) notes that if you have some sparsity (i.e., variables that have no or little impact), the n=10d rule works pretty well… what do we see?
- Also, when the model is fairly smooth, the sample size rule of thumb works pretty well… what do we see?
- Suppose these conditions are violated. Then what?
- Suppose that a computer experiment is run in d-dimensions, but a few are not really active. How should the analysis proceed? Without these dimensions?