Maximum Entropy Sampling for Derivative Information

Noha Youssef
Supervisor: Prof.Henry Wynn

London School of Economics

06 November 2008

Outline

- Introduction to MUCM
- Computer Experiments and Models
- Choosing a Design
- Maximum Entropy Design
- Introduction to Observations Derivatives
- Maximum Entropy Design with Derivatives
- References

Introduction to MUCM

1. My PhD is a part of national project MUCM by EPSRC
2. Managing uncertainty in complex models
3. The aim is to quantify and reduce all sources of uncertainty in the prediction process (Emulators)
4. This includes all types of uncertainty
    - Model inputs (design problem)
    - Model parameters (calibration)
    - Model outputs (Validation)

Computer Models and Experiments

- Complex models (deterministic models): arise from simulators of events or solving equations derived from physical system
- Deterministic simulators are written as complex computer code
- The computer model $y(x)$ used to express the simulator (mimic the behaviour of the simulator)
- Computer experiment: Collection of n runs of a computer model, to investigate $y(x)$ at x_i, $i = 1, \ldots n$
- The design is the collection of n inputs of x

Modelling the Output Computer Experiments

- Gaussian stochastic process used for modelling
- Uses nice properties of the multivariate normal distribution
- Easy to apply the Bayesian approach

$$y(x) = f^T(x)\beta + z(x)$$

- $Y(x)$ is a Gaussian process with mean

$$\mathrm{E}(Y(x)) = X^T\beta$$

and var-cov matrix

$$\mathrm{cov}(Y(x), Y(x)') = \sigma^2 X\Sigma_\beta X^T + \sigma^2\alpha\Sigma_z$$

X design matrix, Σ_β prior var-cov matrix of β, Σ_z var-cov matrix of process $Z(x)$, σ^2 constant, α known rescaling parameter

Choosing the design

- To reduce the uncertainty about the model
  - Reduce inputs uncertainty with clever choice of design points
  - Reduce the error in model structure (calibration-validation)
- Two methods of choosing the design
  - Space filling designs (Latin hypercubes, lattice points, Sobol's sequences)
  - Model-based optimal designs ($D-$, $A-$ optimality,Entropy)

Maximum Entropy Sampling Design (Shewry and Wynn (1987))

- Entropy is negative of information

$$\mathrm{Ent}(Y_s) = \mathrm{E}_{Y_s}[-\log p(Y_s)]$$

  - Y_s vector of observations
  - $p(.)$ density function of Y
  - $D_s = \{x_1, \ldots, x_n\}$
- Partition $Y_N = (Y_S, Y_{N \setminus S})$
- $\mathrm{Ent}(Y_N) = \mathrm{Ent}(Y_S) + \mathrm{E}_{Y_S}\mathrm{Ent}(Y_{N\setminus S}|Y_S)$
- $\mathrm{Ent}(Y_N)$ is fixed
- Min $\mathrm{Ent}(Y_{N\setminus S}|Y_S)$ is $\equiv$ Max $\mathrm{Ent}(Y_S)$
- In the Gaussian case,

$$\mathrm{Ent}(Y_S) = \frac{n}{2}[1 + \log 2\pi] + \frac{\log|\Sigma_s|}{2}$$

- The problem is finding the design that maximizes $\log|\Sigma_s|$

Examples of Covariance Functions

- $\Sigma_z = \sigma^2 \prod_{i=1}^{d} R(x_{it}, x_{is})$
- **Exponential** $R(x_t, x_s) = exp(-\sum_{i=1}^{d} \theta |x_{it} - x_{is}|_i^p)$.
- **Gaussian** $R(x_t, x_s) = exp(-\sum_{i=1}^{d} \theta |x_{it} - x_{is}|^2)$.
- **Brownian Sheet Covariance Matrix**

$$\text{cov}(Z(x_t), Z(x_s)) = \begin{cases} \prod_{i=1}^{d} min(s_i, t_i) & 1 \le i = j \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Getting the Design

- This problem is a combinatorial optimization problem
- Exchange algorithm (fast and efficient)
- Branch and Bound (exact design but not fast(Ko et al,1995))
  - **Idea**
    Dividing the candidate set into subsets
  - **Main Components**
    - Selection of subsets to be explored
    - Bound calculation
    - Branching
  - **Trick**
    Pruning a sub-solution if not improving the current best solution

Examples on MSE

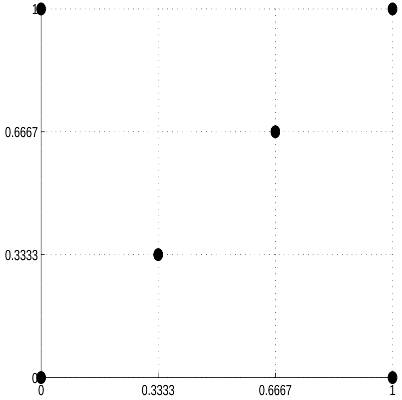


Figure: MSE design of 6-points with $R(x_t, x_s) = exp(-\sum_{i=1}^{d} |x_{it} - x_{is}|^1)$

Sequential Entropy Design

- Splitting the design into many stages
  - Speed up the design process
  - Make use of updated prior information
  - Numerical techniques are required
- Nice formula for sequential entropy found

$$\begin{aligned} \mathrm{Ent}(Y_1, Y_2, ..., Y_n) &= \mathrm{Ent}(Y_1) + E_{Y_1}\mathrm{Ent}(Y_2|Y_1) + E_{Y_2}\mathrm{Ent}(Y_3|Y_2, Y_1) \\ &+ ... + E_{Y_{n-1}}\mathrm{Ent}(Y_n|Y_{n-1}...Y_1) \end{aligned}$$

- Assuming the Gaussian case with known σ^2

$$\mathrm{Ent}(Y_1, \ldots Y_n) = \mathrm{Ent}(Y_1) + \mathrm{Ent}(Y_2|Y_1)$$
$$+\mathrm{Ent}(Y_3|Y_1, Y_2) + \ldots + \mathrm{Ent}(Y_n|Y_1 \ldots Yx_{n-1})$$

- $\mathrm{Ent}(Y_{s \bigcup s'}) = \mathrm{Ent}(Y_s) + \mathrm{E}_{Y_s}\mathrm{Ent}(Y_{s'}|Y_s)$
- max $\mathrm{Ent}(Y_s)$ then max $\mathrm{Ent}(Y_{s \bigcup s'}) - \mathrm{Ent}(Y_s)$
- The updating formula for next point $Y(x_0)$is the determinant of the posterior predictive variance

$$x_0((\sigma_\theta^2\Sigma_\theta)^{-1} + x(\sigma_z^2\Sigma_z)^{-1}x^T)^{-1}x_0^T + \sigma_z^2\Sigma_z$$

Sequential Entropy with unknown σ^2

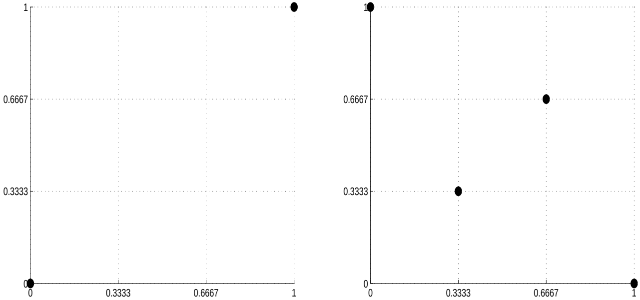
- The whole process $Y(x)$ is a Student t process
- $Y_0(x)|Y(x)$ has Student t distribution
- Entropy of multivariate t is required at each step which is given by

$$-\log \frac{\Gamma(n+p)/2}{(\pi n)^{p/2}\Gamma(n/2)} + \frac{1}{2}\log|\Sigma| + \frac{n+p}{2}\left(\Psi\left(\frac{n+p}{2}\right) - \Psi\left(\frac{n}{2}\right)\right)$$

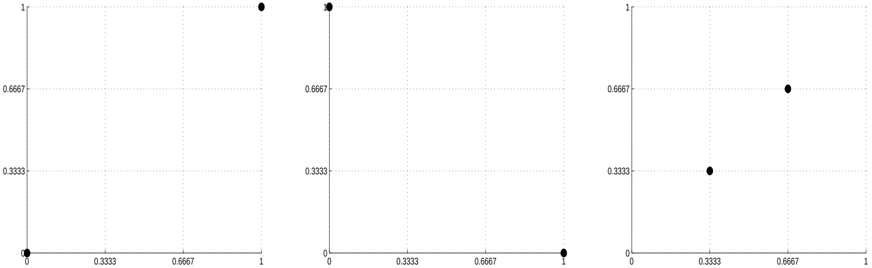
- $\Sigma = V^* a^*$, where

$$\begin{aligned} a^* &= a + \mu^T \Sigma_\theta^{-1} \mu + Y^T \Sigma_z^{-1} Y - (m^*)^T (V^*)^{-1} m^* \\ m^* &= V^*(V^{-1}\mu + x^T \Sigma_z^{-1} y) \\ V^* &= (\Sigma_\theta^{-1} + x^T \Sigma_z^{-1} x)^{-1} \end{aligned}$$

Example (A 6-point MES design obtained over two stages for Exponential covariance function, $\theta = p = 1$



A 6-point MES design obtained over 3 stages for Exponential covariance function, $\theta = p = 1$



What are the derivatives?

- Consider the same model in 2 dimension

$$Y(x) = f^T(x)\theta + Z(x)$$

- Under differentiability condition $Y(x)$ is a Gaussian process $\longrightarrow$ the derivative $\frac{\partial^a}{\partial x_1^{a_1} \partial x_2^{a_2}} Y(x)$ is also Gaussian process
- The prior mean for the derivative

$$E\left(\frac{\partial^a}{\partial x_1^{a_1} \partial x_2^{a_2}} Y(x)|\theta \right) = \frac{\partial^a}{\partial x_1^{a_1} \partial x_2^{a_2}} f^T(x)\theta$$

and the covariance

$$\text{Cov}\left[\frac{\partial^a}{\partial x_1^{a_1} \partial x_2^{a_2}} Y(x), \frac{\partial^b}{\partial x_1^{b_1} \partial x_2^{b^2}} Y(x')|\theta \right] = \frac{\partial^{a+b}}{\partial x_1^a \partial x_2^b} cov(Y(x), Y'(x))$$

- $a_1 + a_2 = a$, $b_1 + b_2 = b$

How would be the model ?

- Instead of just
$$Y_{n\times 1} = (Y(x_1), \ldots, Y(x_n))'$$
we observe
$$Y = (Y(x_1), \ldots Y(x_n), Y^{(1)}(x_1), \ldots, Y^{(1)}(x_n), \ldots, Y^{(k)}(x_1), \ldots, Y^{k}(x_n)$$
of order $n(k+1) \times 1$ where k is the number of derivatives we obtained
- Assuming getting the first derivatives with respect to each variable and subs. in the formula above
- Σ is
$$\begin{pmatrix} \mathrm{var}(Y(x), Y(x)) & \mathrm{cov}(Y(x), Y_1^{(1)}(x)) & \mathrm{cov}(Y(x), Y_2^{(1)}(x)) \\ \mathrm{cov}(Y(x), Y_1^{(1)}(x)) & \mathrm{var}(Y_1^{(1)}(x), Y_1^{(1)}(x)) & \mathrm{cov}(Y_1^{(1)}(x), Y_2^{(1)}(x)) \\ \mathrm{cov}(Y(x), Y_2^{(1)}(x)) & \mathrm{cov}(Y_1^{(1)}(x), Y_2^{(1)}(x)) & \mathrm{var}(Y_1^{(1)}(x), Y_2^{(1)}(x)) \end{pmatrix}$$

The posterior process

- We add those derivatives to our problem, then the posterior mean

$$E(Y_0(x)|Y) = E(Y_0(x)) + (Y(x) - \mu)^T \Sigma^{-1} k(x)$$

  Y includes all the derivatives, $k(x)$ is the vector of covariances of $Y_0(x)$ and $Y(x)$.

- The posterior covariance function is

$$\mathrm{cov}(Y_0(x_t), Y_0(x_s)|Y) = \mathrm{cov}(Y(x_t), Y(x_s)) - k(x_t)^T \Sigma^{-1} k(x_s)$$

Example (The Gaussian covariance function (Nather and Simak (Metrika(2003))))

- $\mathrm{var}(y(x_t), y(x_s)) = \sigma^2 \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$
- $\mathrm{cov}(y(x_t), y_1^{(1)}(x_s)) = \sigma^2 2\theta(x_{1s} - x_{1t}) \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$
- $\mathrm{cov}(y_1^{(1)}(x_t), y(x_s)) = \sigma^2 2\theta(x_{1t} - x_{1s}) \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$
- $\mathrm{cov}(y_1^{(1)}(x_t), y_1^{(1)}(x_s)) =$
  $\sigma^2(2\theta - 4\theta^2(x_{1s} - x_{1t})^2) \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$
- $\mathrm{cov}(y_2^{(1)}(x_t), y_2^{(1)}(x_s)) =$
  $\sigma^2(2\theta - 4\theta^2(x_{2s} - x_{2t})^2) \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$
- $\mathrm{cov}(y_1^{(1)}(x_t), y_2^{(1)}(x_s)) =$
  $\sigma^2(2\theta - 4\theta^2(x_{1t} - x_{1s})(x_{2s} - x_{2t})) \Sigma_{i=1}^2 \exp(-\theta(x_{it} - x_{is})^2)$

Entropy Design Using Derivatives

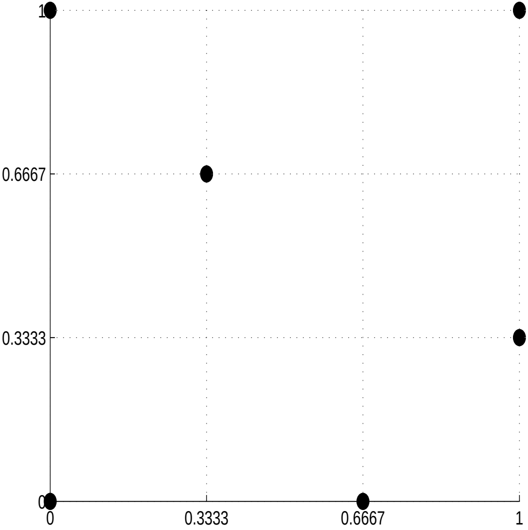
- If all derivatives are taken into account, to find an n point design we max

$$\mathrm{Ent}(Y(x_1), \ldots Y(x_n), Y_1^{(1)}(x_1), \ldots Y_1^{(1)}(x_n), Y_2^{(1)}(x_1), \ldots Y_2^{(1)}(x_n)$$

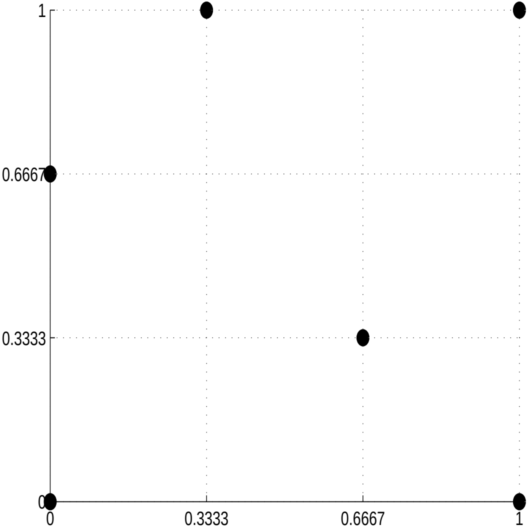
- We could also obtain $\mathrm{Ent}(Y_s)$

$$\mathrm{Ent}(Y_s, Y_s^{*}) = \mathrm{Ent}(Y_s) + \mathrm{EEnt}(Y_s^{(*)}|Y_s)$$

Example (A 6-point MES design without using derivatives)



Example(A 6 point MES design using observations and derivatives)



Why using observations derivatives (Morris *et al* 1993)

- Many simulators arise from solving differential equations
- Some simulators require the gradient to be approximated by a difference equation
- Derivatives are used to identify the inputs that have the greatest or least effect on the response
- Sometimes they reduce the computational expenses or improving the process of getting the emulators
- They are beneficial on modeling nonlinear and dynamic systems
- In case the derivatives don't exist, they can be produced by
  - Divided difference approach
  - System of adjoint equations ⇢ sensitivity analysis
  - Automatic differentiation

Conclusions and Future research

- In case of entropy of t distribution,
  - The updating formula needs experimenting data every stage
  - Expectation needed every stage
- Using the sequential approach with derivatives

References

- Gelman, A.et al (2004). Bayesian data analysis, (second ed.). Texts in Statistical Science Series. Chapman and Hall/CRC, Boca Raton, FL.
- Hoffman et al, IBM Research Report, (2000)
- Ko et al, IBM Research Report, (1995)
- Koehler and Owen, Handbook of Statist.,(1996)
- Morris M. et al,Technometrics,(1993)
- Sebastiani and Wynn, J. R. Stat. Soc. Ser. B Stat. Methodol., (2000)
- Shewry and Wynn , J. of Apply Statistics,(1977)