Quantifying Uncertainty

Sai Ravela

M. I. T

Last Updated: Spring 2013

Typical Problems

- What happened in past, "climate", "economy", "terrorism","evolution"
- What is the future of "weather," "climate", "economy"...
- How good are my models?
- What is a good model?

These are fundamental challenges of all time, but climate related issues are front and center now.

Objectives

- Predictions demand a quantification of uncertainty
- Estimates demand a quantification of Uncertainty
- Inference demands a quantification of Uncertainty

Quantifying Uncertainty

- Uncertainty can be Aleatory: a random outcome.
- Uncertainty can be Epistemic: an unknown quantity to be estimated or imputed.

Systems Perspective

- Uncertainty in Model Predictions: State Estimation
- Uncertainty in Model Parameters: Parameter Estimation
- Uncertainty in Model Structure: Model Selection

These are most easily seen in time-dependent processes, though it is just as valid for statistical predictive and emperical modeling.

It might take both

This is a dilemma for modeling many physical processes:

1. Physics-based models applicable to the full-range of dynamics, but difficult to implement and often with too-many degrees of freedom for the problem of interest.
2. Empirical ones can't generalize, limited predictability.

Often, it takes both skills to build a good model but the two don't speak the same language or communicate well.
As Walker says:
There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only without knowledge of each other's work, but also will ignore the problems which require mutual assistance.

Main Areas

- Linear Models and Gaussian Inference
- Two-point Boundary Value Problems
- Filters and Smoothers
- Ensemble Methods
- Sampling and Markov Chain Monte-Carlo
- Hierarchical Bayes

This was a lot of material (but there is more to cover).

Central Limit, a good start?

- The sampling distribution of the statistic can be estimated by repeatedly drawing n-length sample sequences from a distribution, calculating the statistic and then considering the resulting distribution.
- If the original distribution had a mean m and variance s, then in large n, the sampling distribution:
- Converges to a Gaussian.
- The sample mean converges $m_s \to m$ and has variance $v = \frac{s}{n}$
- The sample variance converges as $(n-1)v = \chi^2(n-1)$.
- Small sample problem!

How to Solve?

Least Squares:

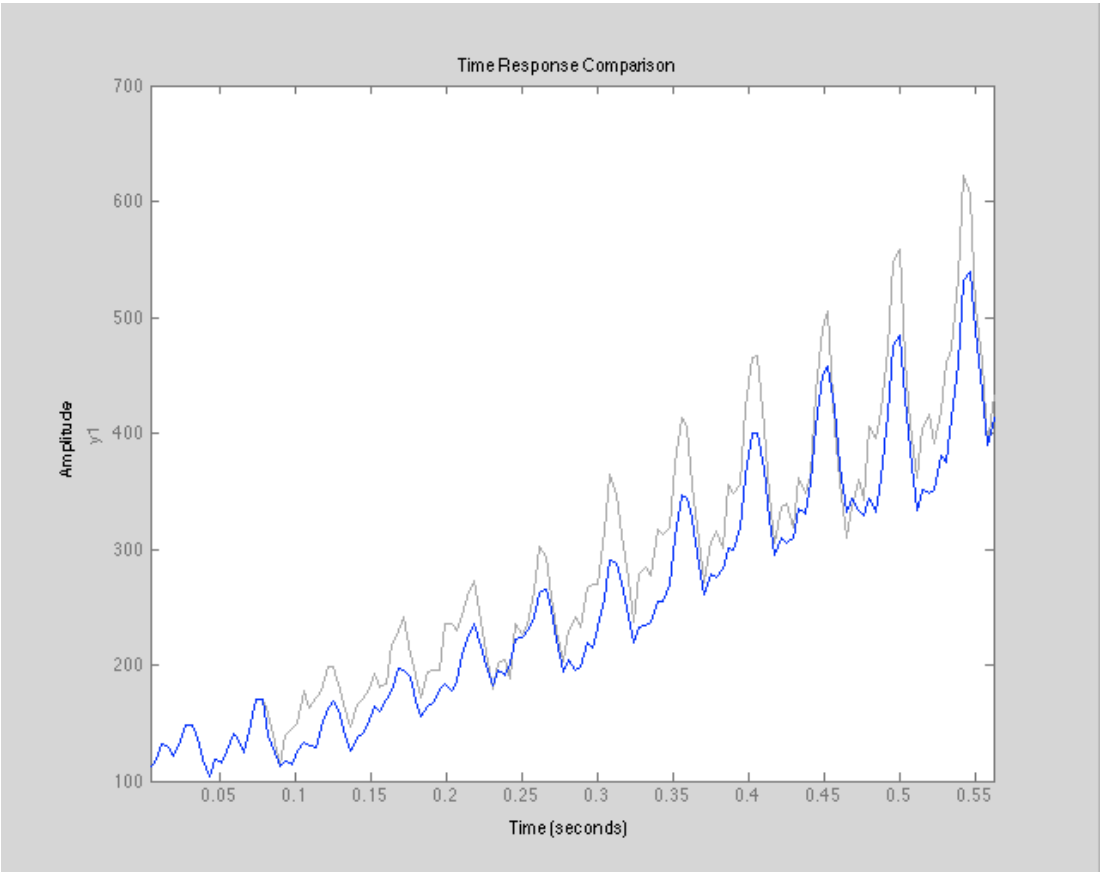
$$J(\underline{\alpha}) := ||\underline{\mathrm{x}} - H\underline{\alpha}||$$

Understand the notation and terms.

$$\begin{aligned} dJ/d\underline{\alpha} &= 0 && (1)\\ \Rightarrow\ & H^T\underline{x} = H^T H\underline{\alpha} && (2)\\ \Rightarrow\ & \underline{\hat{\alpha}} = (H^T H)^{-1} H^T \underline{\mathrm{x}} && (3) \end{aligned}$$

Least Squares Estimate using the Pseudo inverse. Stationary Point.

AR model identification



Bayesian Estimate

$$P(\underline{\alpha}|\underline{x}) \propto P(\underline{x}|\underline{\alpha})P(\underline{\alpha}) \tag{4}$$

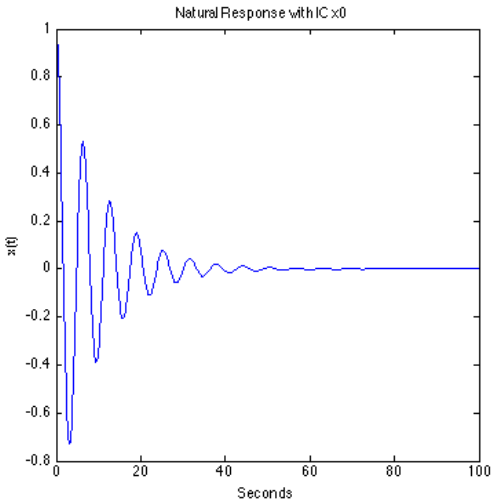
Maximum a posteriori (Bayes) estimate, when $\underline{\eta} \sim N(0, C_{XX})$ and $\underline{\alpha} \sim N(\underline{\bar{\alpha}}, C_{\alpha\alpha})$, of mean and covariance:

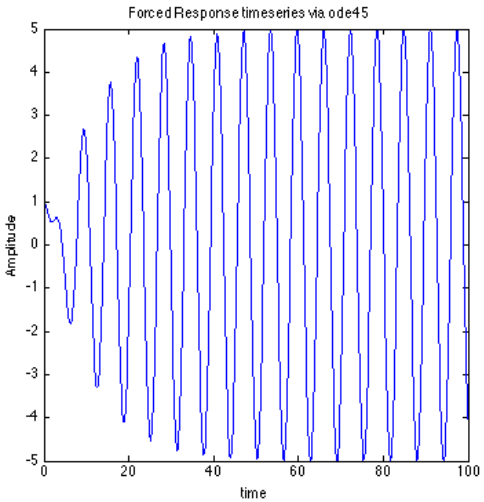
$$\begin{aligned}
\underline{\hat{\alpha}} &= \underline{\bar{\alpha}} + C_{\alpha\alpha}H^T(HC_{\alpha\alpha}H^T + C_{xx})^{-1}(\underline{x} - H\underline{\bar{\alpha}}) \\
\hat{C}_{\alpha\alpha} &= (H^T C_{xx}^{-1} H + C_{\alpha\alpha}^{-1})^{-1} \\
&= C_{\alpha\alpha} - C_{\alpha\alpha}H^T(HC_{\alpha\alpha}H^T + C_{xx})^{-1}HC_{\alpha\alpha}
\end{aligned}$$

Chalk talk: HOW IS THIS DERIVED?

Spring Mass System

$$\dot{X} = AX + F \tag{5}$$





Discretization

We can solve

$$\dot{X} = AX + F \tag{6}$$

using numerical methods, e.g. Runge-Kutta methods (In matlab, ode45). Let's do demo.
But to understand, let's take an Euler discretization with zero-order hold of forcing. Then

$$A_d = e^{A\Delta t} = \mathcal{L}^{-1}[(sI - A)^{-1}]_{t=\Delta t} \tag{7}$$

$$F_d = A^{-1}(A_d - I)B \tag{8}$$

$$x_{t+\Delta t} = A_d x_t + F_d \tag{9}$$

We Just Say

$$x_{n+1} = M(x_n; \underline{\alpha}) \tag{10}$$

Time Dependent Example

$$\underline{x}_{n+1} = M(\underline{x}_n; \underline{\alpha})$$
$$\underline{y}_n = H\underline{x}_n + \underline{\eta}$$

We have assumed the parameter vector is known constant, the model is deterministic, the observations are linearly related, but additively noisy and time-independent with $\underline{\eta} = N(\underline{0}, R)$. We are given a series of measurements $\underline{y}_0 \ldots \underline{y}_m$ and we are asked to estimate the initial condition $\underline{x}_0$. We may simply produce a least-squares function:

$$J(\underline{x}_0) := (\underline{x}_0 - \underline{x}_b)^T C_{00}^{-1} (\underline{x}_0 - \underline{x}_b) + \sum_{i=1}^{m} (\underline{\mathrm{y}}_i - H\underline{\mathrm{x}}_i)^T R^{-1} (\underline{\mathrm{y}}_i - H\underline{\mathrm{x}}_i)$$

"Forward Backward"

Forward(from $\underline{\lambda}$)

$$\begin{aligned} \underline{x}_i &= M(\underline{x}_{i-1}; \underline{\alpha}) \\ & 0 < i < M \end{aligned}$$

Backward(from $\underline{x}$)

$$\begin{aligned} \underline{\lambda}_m &= H^T R^{-1}(\underline{y}_m - H\underline{x}_m) \\ \underline{\lambda}_i &= \frac{\partial M^T}{\partial x_i}\underline{\lambda}_{i+1} + H^T R^{-1}(\underline{y}_i - H\underline{x}_i) \\ \underline{\hat{x}}_0 &= \underline{x}_b + C_{00}\frac{\partial M^T}{\partial x_0}\underline{\lambda}_1 \end{aligned}$$

Uncertainty?

Via Linearization

Forward(you'll need this in the end)

$$C_{ii} = \frac{\partial M}{\partial x_{i-1}} C_{i-1i-1} \frac{\partial M^T}{\partial x_{i-1}} \; 0 < i \leq m$$

What about backward? Convenient via information form:

$$\begin{aligned}
\hat{I}_{mm} &= H^T R^{-1} H \\
\hat{I}_{ii} &= \frac{\partial M^T}{\partial x_i} I_{i+1i+1} \frac{\partial M}{\partial x_i} + H^T R^{-1} H \\
\hat{C}_{00} &= \left[C_{00}^{-1} + \frac{\partial M}{\partial x_0} \hat{I}_{11} \frac{\partial M^T}{\partial x_0} \right]^{-1}
\end{aligned}$$

Here $\mathcal{L} = \frac{\partial M}{\partial x_i}$ is the Jacobian of M and $\mathcal{L}^T$ is its adjoint.

Example: Double Pendulum

With a measurement every second with uncertainty same as C_{00}

$$\hat{C}_{00} = \text{1.0e-03 } * \begin{matrix} 0.0130 & -0.0033 & 0.0127 & -0.0009 \\ -0.0033 & 0.1259 & -0.0004 & -0.0148 \\ 0.0127 & -0.0004 & 0.0268 & -0.0061 \\ -0.0009 & -0.0148 & -0.0061 & 0.2176 \end{matrix}$$

The Missing Data Estimate and Uncertainty

Estimate:

$$M = C_{yx} C_{xx}^{-1} \tag{11}$$

$$\hat{\underline{y}} = C_{yx} C_{xx}^{-1} \hat{\underline{x}} \tag{12}$$

$$\tag{13}$$

Uncertainty follows from objective:

$$\begin{aligned} J(\underline{x}_i) := & (\underline{y}_i - M\underline{x}_i)^T C_{yy}^{-1} (\underline{y}_i - M\underline{x}_i) \\ & + (\underline{x}_i - \bar{\underline{x}}_i)^T C_{xx}^{-1} (\underline{x}_i - \bar{\underline{x}}_i) \end{aligned} \tag{14}$$

Fisher Information

Uncertainty Estimate:

$$d^2 J / d\underline{x_i}^2 = M^T C_{yy}^{-1} M + C_{xx}^{-} 1 \quad (15)$$

$$\hat{C}_{yy} = (C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} C_{xx}^{-1} + C_{xx}^{-1})^{-1} \quad (16)$$

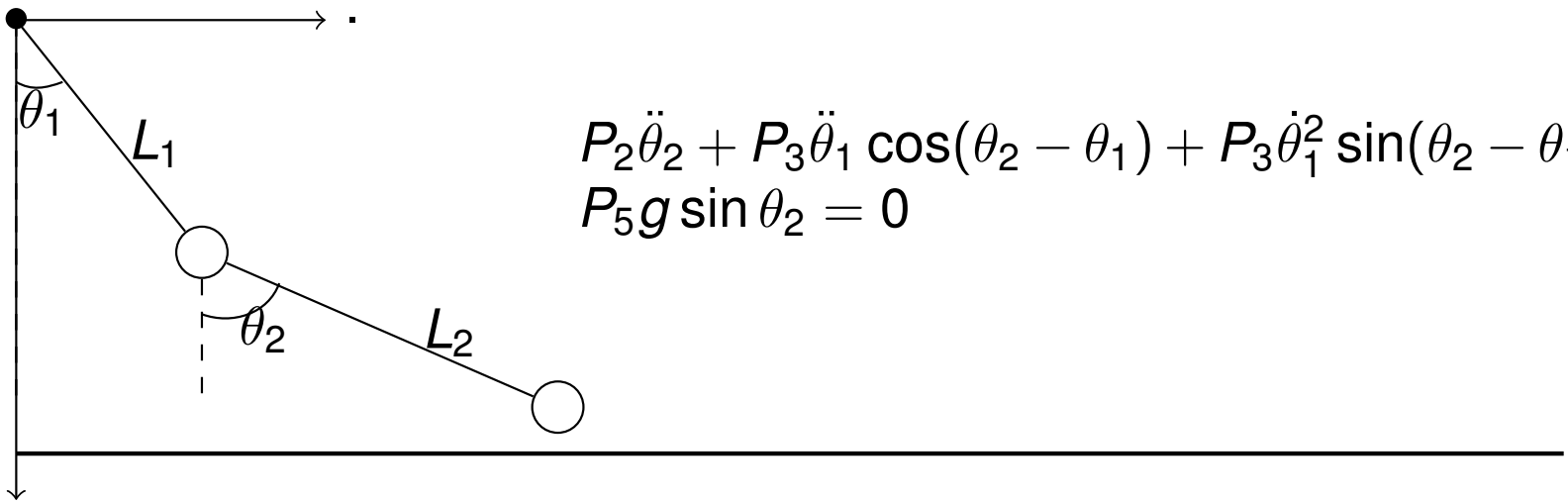
$$(17)$$

Uncertainty follows from objective:

$$\begin{aligned} J(\underline{x}_i) := & (\underline{y_i} - M\underline{x_i})^T C_{yy}^{-1} (\underline{y_i} - M\underline{x_i}) \\ & + (\underline{x_i} - \underline{\bar{x}_i})^T C_{xx}^{-1} (\underline{x_i} - \underline{\bar{x}_i}) \end{aligned} \quad (18)$$

Towards a Nonlinear World.

$$P_1\ddot{\theta}_1 + P_3\ddot{\theta}_2\cos(\theta_2-\theta_1) - P_3\dot{\theta}_2^2\sin(\theta_2-\theta_1) + P_4 g\sin\theta_1 = 0$$

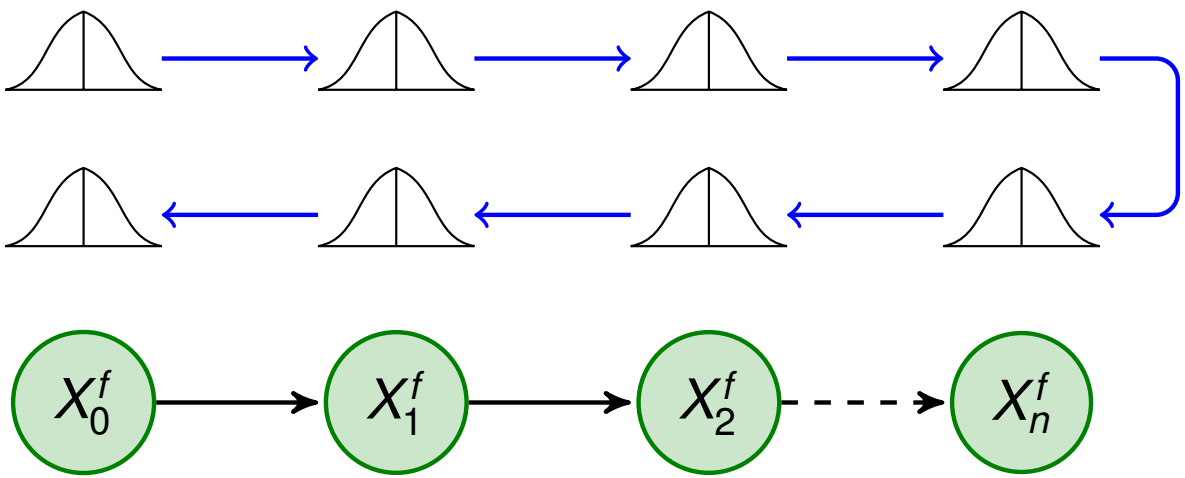


$$P_2\ddot{\theta}_2 + P_3\ddot{\theta}_1\cos(\theta_2-\theta_1) + P_3\dot{\theta}_1^2\sin(\theta_2-\theta_1) + P_5 g\sin\theta_2 = 0$$

Let,

$$P_1 = (m_1+m_2)L_1^2 \qquad P_4 = (m_1+m_2)L_1$$
$$P_2 = m_2L_2^2 \qquad P_5 = m_2L_2$$
$$P_3 = m_2L_1L_2$$

Uncertainty Propagates in Time-Depdendent Processes



$$M \equiv M(\underline{x}_n; \alpha_n)$$

$$\underline{x}_{n+1} = M(\underline{x}_n; \alpha_n) + \underline{\omega}_n$$

M: -Physical or Statistical Model

Filters and Smoothers

Sequential Filtering:

$$P(\underline{x}_n|\underline{y}_1 \dots \underline{y}_n) \propto P(\underline{y}_n|\underline{x}_n) \sum_{x_{n-1}} P(\underline{x}_n|\underline{x}_{n-1}) P(\underline{x}_{n-1}|\underline{y}_1 \dots \underline{y}_{n-1}) \quad (19)$$

$$= P(\underline{y}_n|\underline{x}_n) P(\underline{x}_n|\underline{y}_1 \dots \underline{y}_{n-1}) \quad (20)$$

$$= P(\underline{y}_n|\underline{x}_n) P(\underline{x}_n^f) \quad (21)$$

The recursive form is simple when a perfect model is assumed, but

the Kolmogorov-Chapman equation has to be used in the presence of model error. $P(\underline{x}_n|\underline{y}_1 \dots \underline{y}_{n-1})$ is the forecast distribution or prior distribution also seen as $P(\underline{x}_n^f)$

Write the Objective

Sequential Filtering:

$$J(\underline{x}_n) := \frac{1}{2}(\underline{x}_n - \underline{x}_n^f)^T P_f^{-1}(\underline{x}_n - \underline{x}_n^f) + \frac{1}{2}(\underline{y}_n - H\underline{x}_n)^T R^{-1}(\underline{y}_n - H\underline{x}_n) \quad (22)$$

We have assumed a linear observation operator $\underline{y}_n = H\underline{x}_n + \underline{\eta}$, with $\eta \sim N(0, R)$.

Find the Stationary Point

Sequential Filtering:

$$\begin{aligned}
\hat{\underline{x}}_n &= \underline{x}_n^f + P_f H^T (H P_f H^T + R)^{-1} (\underline{y}_n - H \underline{x}_n^f) && (23)\\
&= \underline{x}^a && (24)\\
P_a &= (H^T R^{-1} H + P_f^{-1})^{-1} && (25)\\
&= P_f - P_f H^T (H P_f H^T + R)^{-1} H P_f && (26)
\end{aligned}$$

Then, launch a new prediction $\underline{x}_{n+1}^f = M(\underline{x}_n)$ and the new uncertainty (predicted) is $P_f = L P_a L^T$, where $L = \frac{\partial M}{\partial \underline{x}_n}$ when the model is nonlinear. Propagating produces the moments of $P(\underline{x}_{n+1} | \underline{y}_1 \ldots \underline{y}_n)$.

Smoother

We are interested in the state estimates at all points in an interval, that is:

$$P(\underline{x}_1 \dots \underline{x}_n | \underline{y}_1 \dots \underline{y}_n) \tag{27}$$

The joint distribution can account for model errors, state and parameter errors within its framework.
We break it down via Bayes Rule, Conditional Independence and Markov assumption, and marginalization and perfect modelassumption, leading to a coupled set of equations that are recursively solved.

Fast Calculation

$$A^a = A^f + \tilde{A}^f \tilde{\Omega}^{fT} [US^{-2}U^T][Z - \Omega^f]$$

$$(n,s) \quad (n,s) \quad (n,s)(s,n)(n,s)(s,s)(s,n)(n,s) \quad (n,s)$$

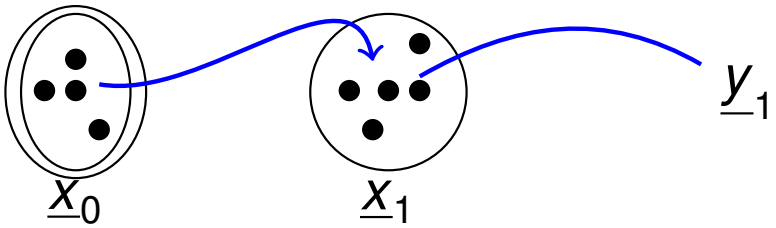
Return by right to left, multiply; FAST, low-dimensional

$$\begin{aligned} A^a &= A^f + \tilde{A}^f X_5 \\ &= A^f (I_s + X_4) \\ &= A^f X_5 \end{aligned}$$

A “weakly” nonlinear transformation ($X_5 \equiv X_5(A^f)$)

Plug and Play

So,

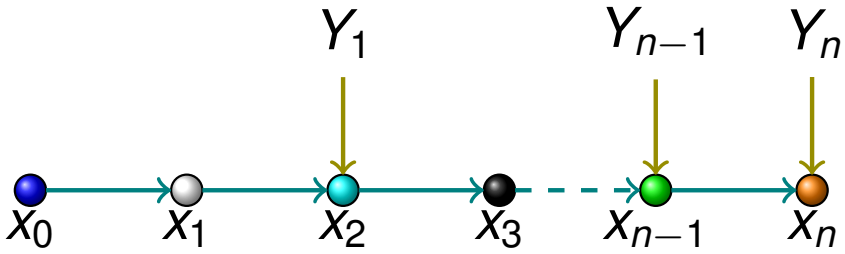


$$A_0^a = A_0^f I_s \leftarrow \text{No measurement}$$
$$A_1^a = A_1^f X5_1 \leftarrow \text{Filter, same as } X_5$$
$$A_1^s = A_1^a I_s \leftarrow \text{No future measurement}$$
$$A_0^s = A_0^a + \tilde{A}_0^a \tilde{\Omega}_1^{fT} [U_1 S_1^{-2} U_1^T][Z_1 - \Omega_1^f]$$
$$= A_0^a X5_1$$

Note: $X5$ here is same as X_5 in earlier slide.

Fixed Interval & Fixed Lag

Fixed Interval

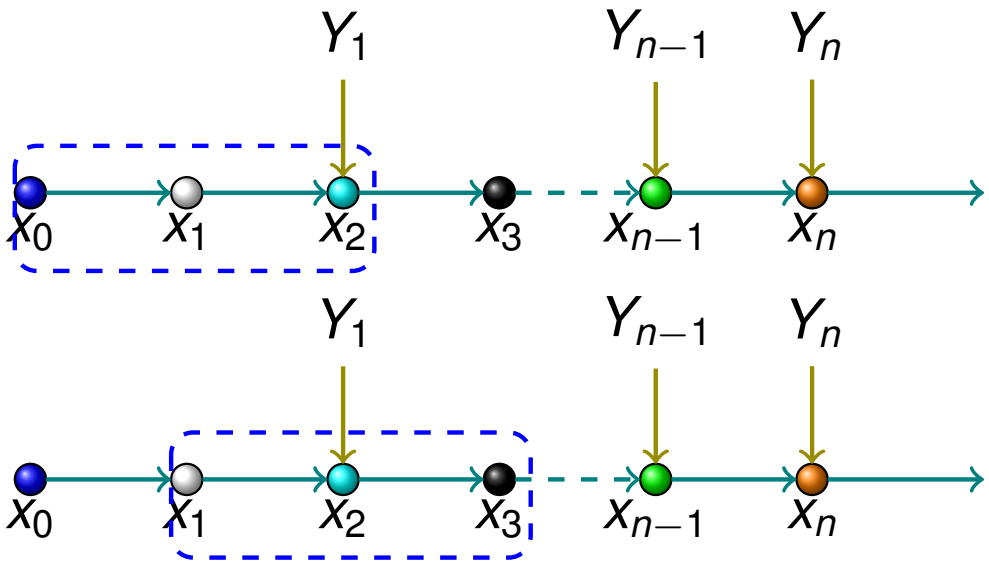


$$\left.\begin{array}{l} P(\underline{x}_0|\underline{y}_1 \cdots \underline{y}_n) \\ P(\underline{x}_1|\underline{y}_1 \cdots Y_n) \\ \vdots \end{array}\right\} \textit{Smoother}$$

$$P(\underline{x}_n|\underline{y}_1 \cdots \underline{y}_n) \leftarrow \textit{Filter}$$

Fixed Interval & Fixed Lag

Fixed Lag



$$P(\underline{x}_0|\underline{y}_1 \ldots \underline{y}_n) \cong P(\underline{x}_0|\underline{y}_1 \ldots \underline{y}_L), \quad (L < n)$$

$$P(\underline{x}_i|\, Y_0 \ldots \underline{y}_{i+L})$$

Smothed up to a “window”

Backward Recursion: Fast Ensemble Smoothing

Key Assumption: Jointly Gaussian Distributions.

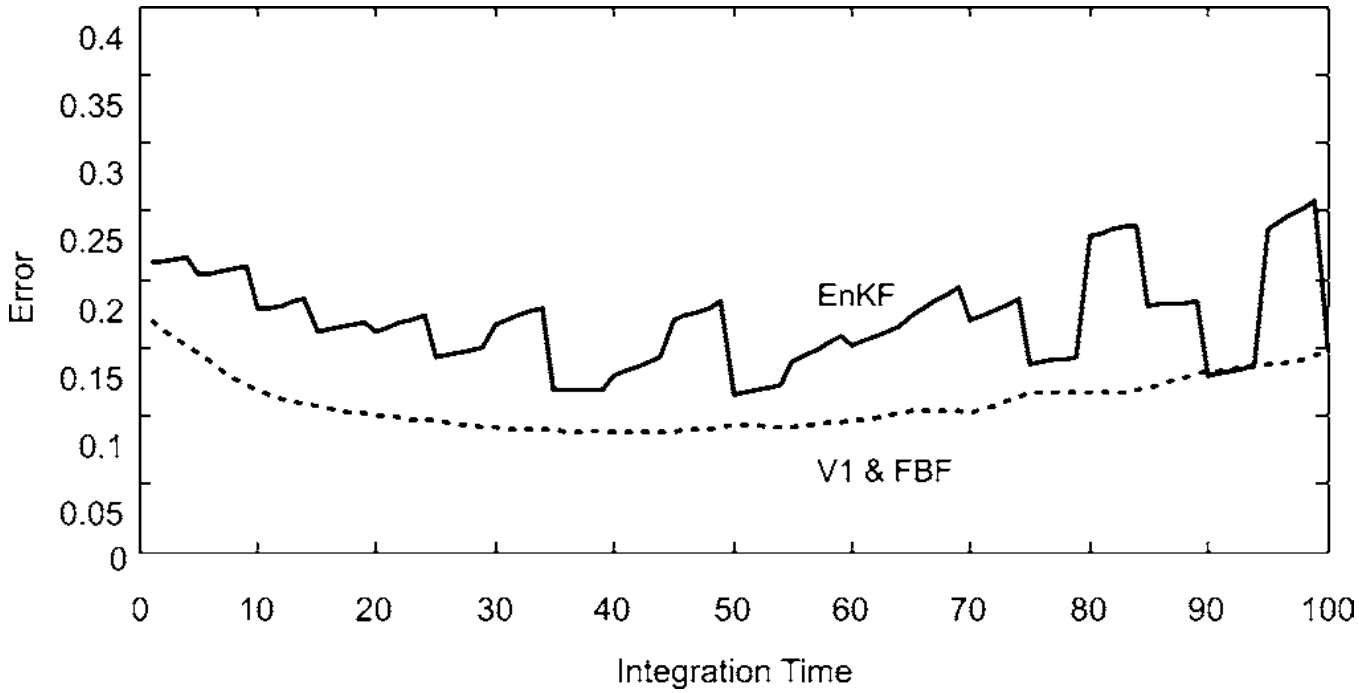
$$A_k^s = A_k^a \prod_{j=k+1}^{N} X5_j$$

$$C_k = \prod_{j=k+1}^{N} X5_j = X5_{k+1} C_{k+1}$$

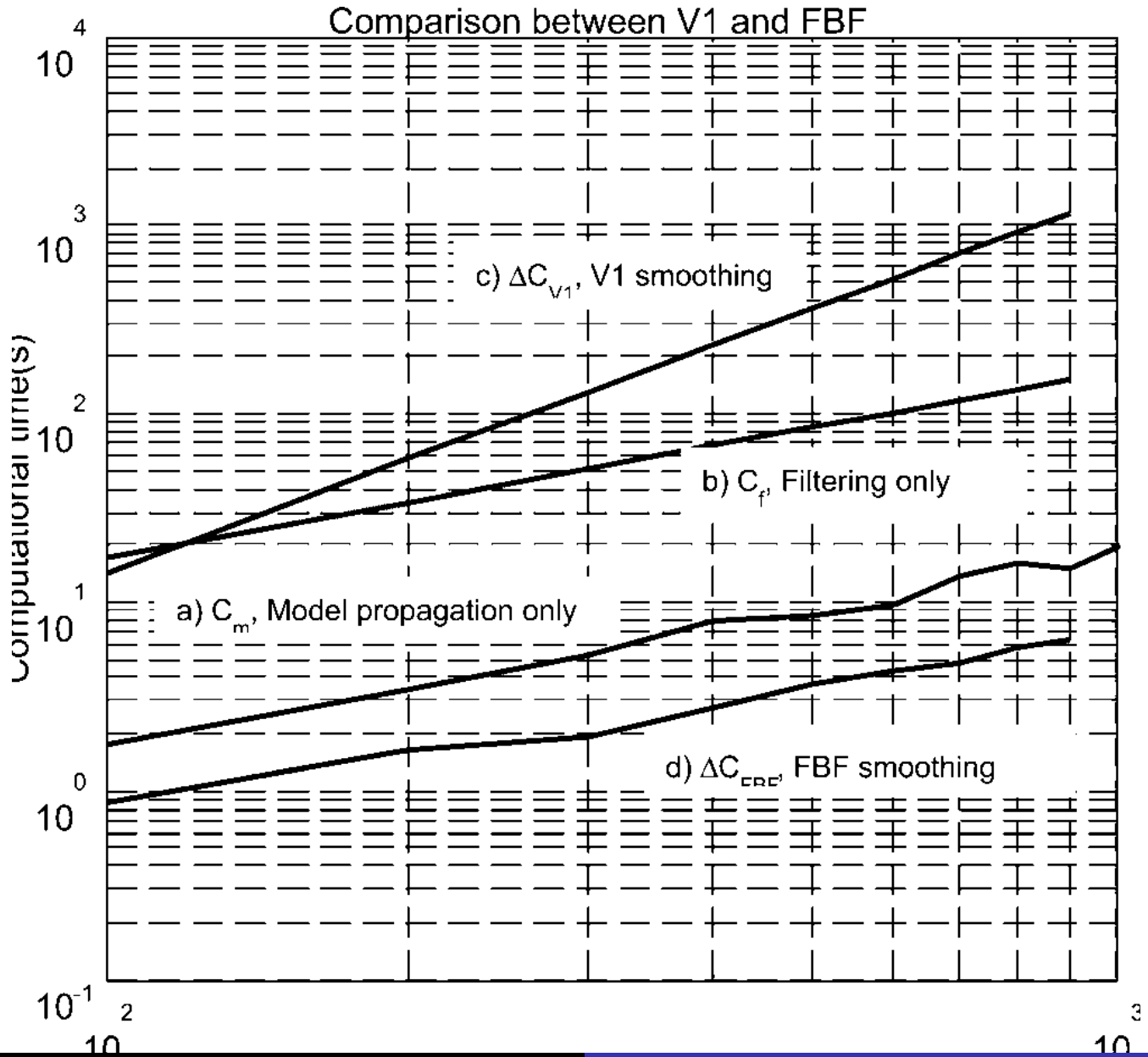
Fixed Lag is FIFO

$$
\begin{aligned}
A_k^s &= A_k^a \prod_{j=k+1}^{k+w} X5_j \\
&= A_k^a C_k \\
C_k &= X5_k^{-1} C_{k-1} X5_{k+w}
\end{aligned}
$$

Fixed Interval on Lorenz

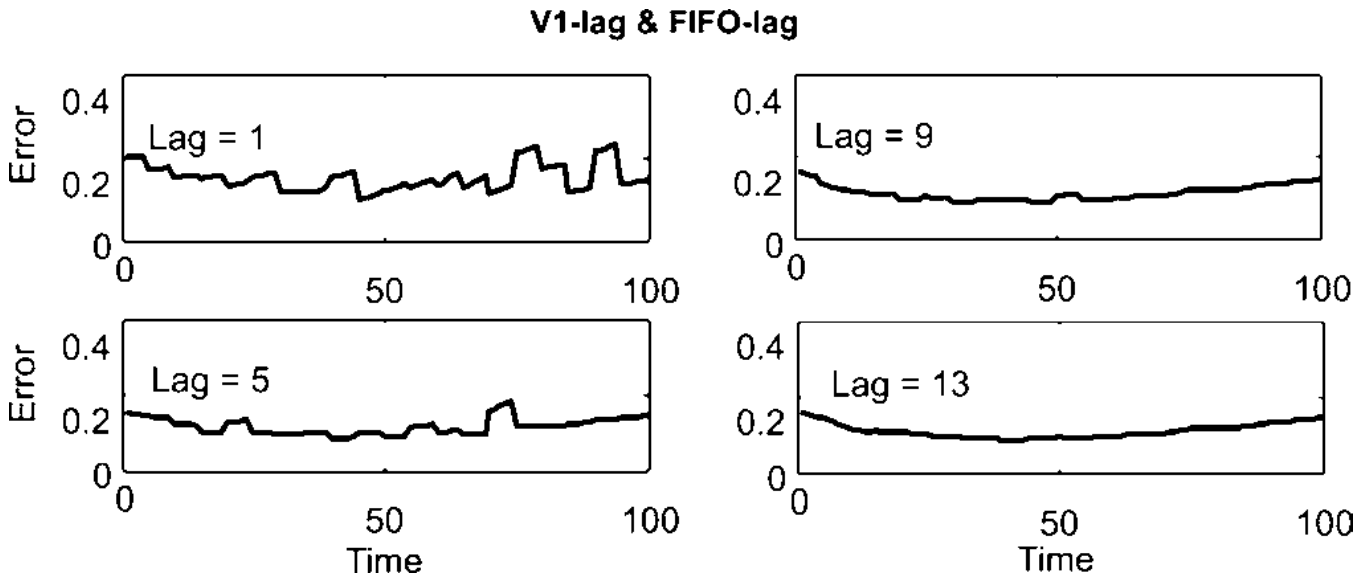


Costs of Inference, Toy Problem



Fixed Lag

Fixed Lag Smoother



Ways to simplify Models for Uncertainty Propagation

1. Spectral Truncation: Find a few leading directions of Covariance or Model and propagate them. Breed Vectors. Calculate a reduced local linear model from ensemble.
2. Localization: Localize filtering and smoothing, use scale-recursive decomposition.
3. Model Reduction: Reduce order of linearized model, construct a reduced model from snapshots.
4. Sample Input-Output pairs to create a simple auxiliary model.

SV Ensemble

Now,let C_1 be a metric on vector $\underline{u}_1$ and let C_0 be a metric on $\underline{u}_0$

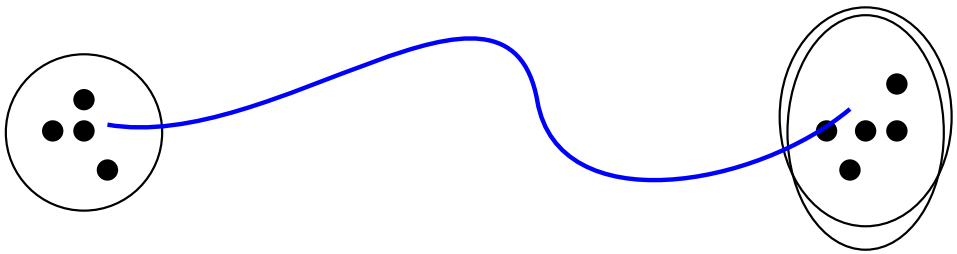
$$\lambda = \frac{< \mathcal{L}\underline{u}_0, C_1 \mathcal{L}\underline{u}_0 >}{< \underline{u}_0, C_0 \underline{u}_0 >} = \frac{< \underline{u}_0, \mathcal{L}^{\#} C_1 \mathcal{L}\underline{u}_0 >}{< \underline{u}_0, C_0 \underline{u}_0 >}$$

Maximize ratio for the k^{th} perturbation: λ_k:

$$\Rightarrow \mathcal{L}^{\#} C_1 \mathcal{L}\underline{u}_0^{(k)} = \lambda_k C_0 \underline{u}_0^{(k)}$$

Which is a generalized eigenvalue problem. Note that when $C_1 = I$, and $C_0 = P_0^f$ then $\underline{u}_1^{(k)}$ are leading directions of P_1^f

Breeding



Initial
Perturbation

Align with leading
directions of error growth
(Lyapunov vectors)

$$\underbrace{Q_{i+1}R_{i+1}}_{\text{Q R decomposition}} = \underbrace{\mathcal{L}}_{TLM} Q_i$$

$$Q_0 \equiv I \qquad Q_0 \rightarrow Q_i \cdots \underbrace{Q_k}_{forgets Q_0}$$

Alternate form

- Let $X = USV^T$ be the singular value decomposition and here $S_{ii} \geq S_{i+1i+1}$. Then $D = X^T X = V \Lambda V^T$ where $\Lambda = S^2$, a small matrix.
- We calculate the eigen vectors and eigen values of D recursively. Let $D_1 = D$; and for $k = 1 \ldots d$

$$\underline{v}_k = PowerIteration(D_k) \quad (28)$$

$$\lambda_{kk} = \underline{v}_k^T D_k \underline{v}_k \quad (29)$$

$$D_{k+1} = D_k - \underline{v}_k \lambda_{kk} \underline{v}_k^T \quad (30)$$

Noting that $S_d = \sqrt{\Lambda_d}$, we obtain U_d as a skinny *nxd* matrix:

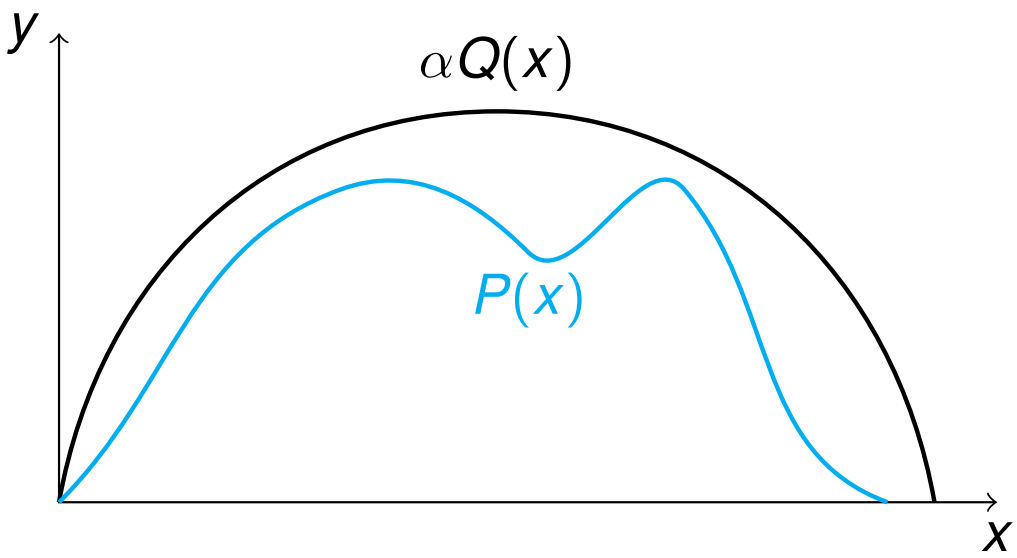
$$U_d = X V_d S_d^{-1} \quad (31)$$

Store U_d and Λ_d and use them to calculate the norm in an application. DEMO IN MATLAB

Markov Chain Monte Carlo

- Monte Carlo sampling made for large scale problems via Markov Chains
  - Monte Carlo Sampling
  - Rejection Sampling
  - Importance Sampling
  - Metropolis Hastings
  - Gibbs
- Useful for MAP and MLE problems

Rejection Sampling



$$\alpha Q(x) \geq P(x)$$

$$x_i \sim Q(x), \quad y_i \sim U[0, \alpha Q(x_i)]$$

Importance Sampling

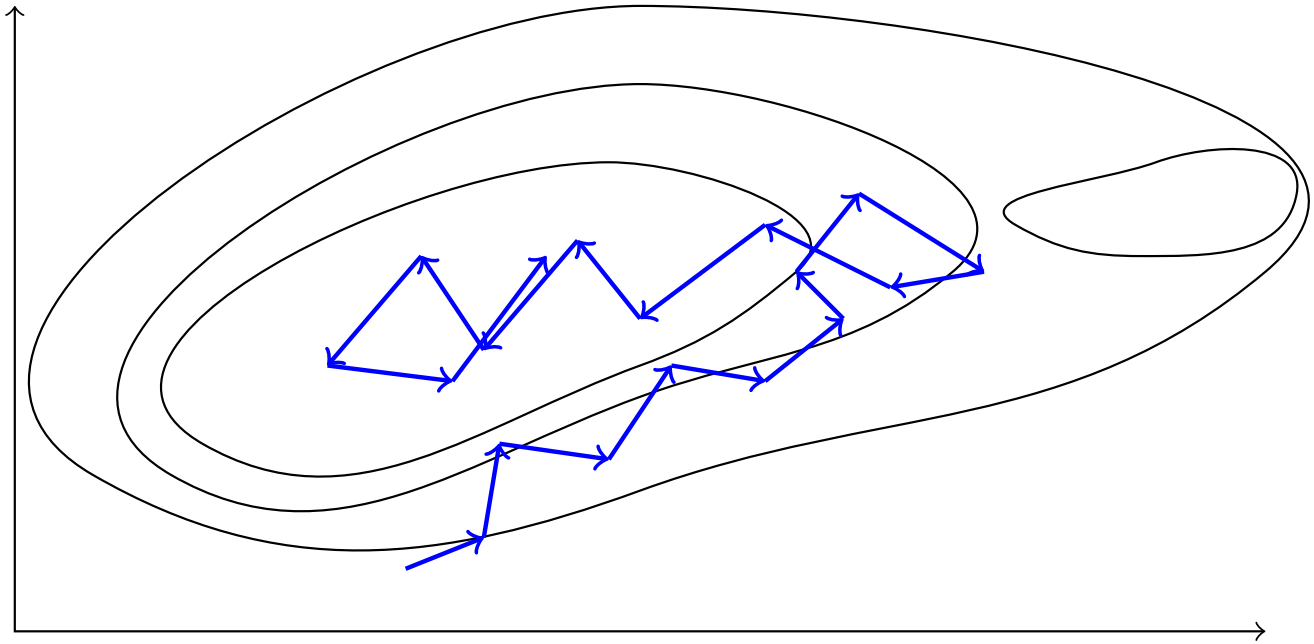
$$\int f(x)P(x)dx = \int f(x)\frac{P(x)}{Q(x)}Q(x)dx$$

$$\cong \frac{1}{S}\sum_{s=1}^{S} f(x_s)\frac{P(x_s)}{Q(x_s)}, \quad x_S \sim Q(x)$$

$$\frac{P(x_s)}{Q(x_s)} \equiv \text{Importance of sample} \doteq \omega_s$$

$$\hat{I}_S = \frac{1}{S}\sum_{s=1}^{S} f(x_s)\omega_s$$

Markov Chain Monte Carlo



1. A proposal distribution from local moves (not globally, as in RS/IS).
   1.1 Local moves could be in some subspace of state space.
2. Move is conditioned on most recent sample

Metropolis Hastings

Draw $x' \sim Q(x'; x)$, the proposal distribution

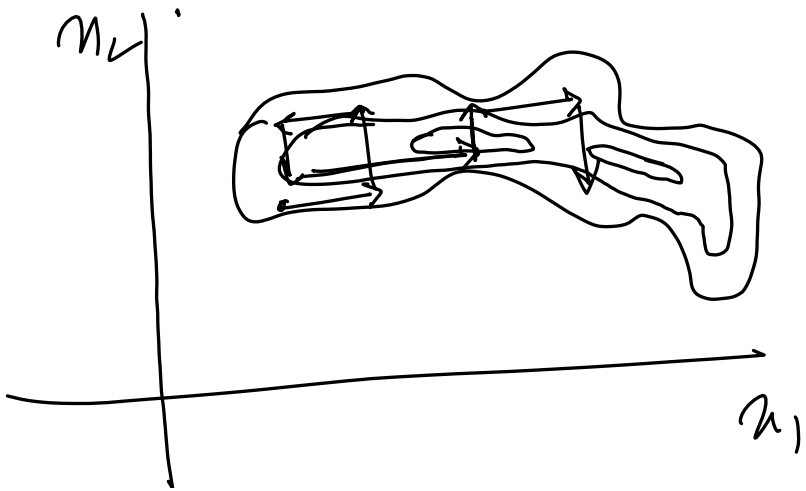
$$a = \min\left(1, \frac{P(x')Q(x; x')}{P(x)Q(x'; x)}\right)$$

Accept x' with prob. a, else retain x.

- $\Rightarrow$ No need to have pmf in $Q(x'; x)$
- $\Rightarrow$ Satisfies detailed balance
- $\Rightarrow$ Equilibrium distribution is target distribution

Note: $P_T(x \to x') = aQ(x'; x)$

Gibbs Sampler: a different transition



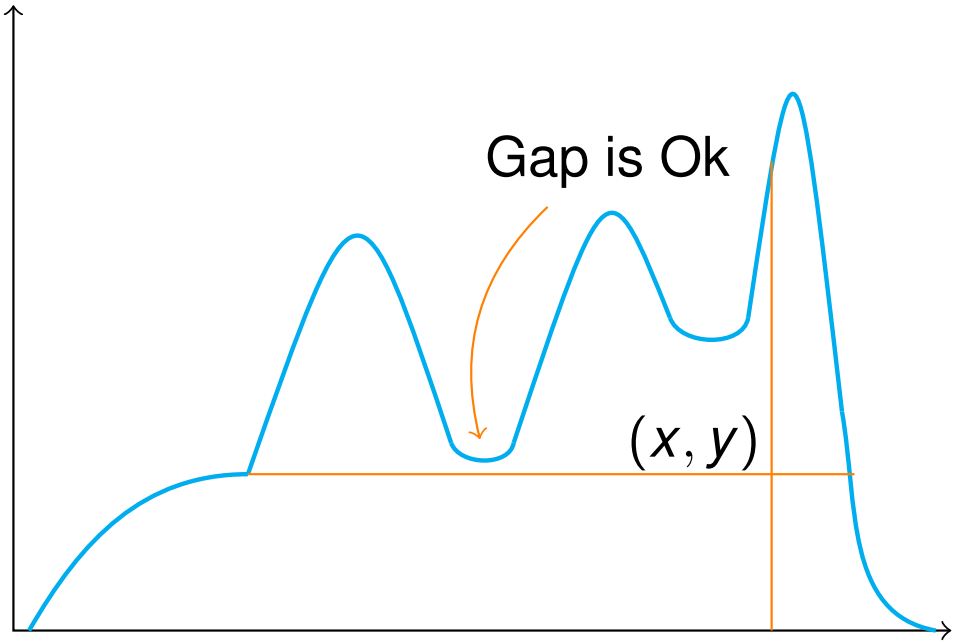
Let $\underline{x} = x_1, \cdots, x_n$
(a huge dimensional space) and we want to sample

$$P(\underline{x}) = P(x_1 \cdots x_n)$$
$$P(\underline{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1) \ldots P(x_n|x_{n-1} \ldots x_1)$$

Gibbs:

$$P(x_1) \rightarrow P(x_2|x_1) \rightarrow P(x_3|x_1, x_2) \rightarrow \cdots$$
$$\rightarrow P(x_n|x_n - 1 \ldots x_1) \rightarrow P(x_1|x_{i\neq 1}) \rightarrow P(x_2|x_{i\neq 2}) \ldots$$

Slice Sampler

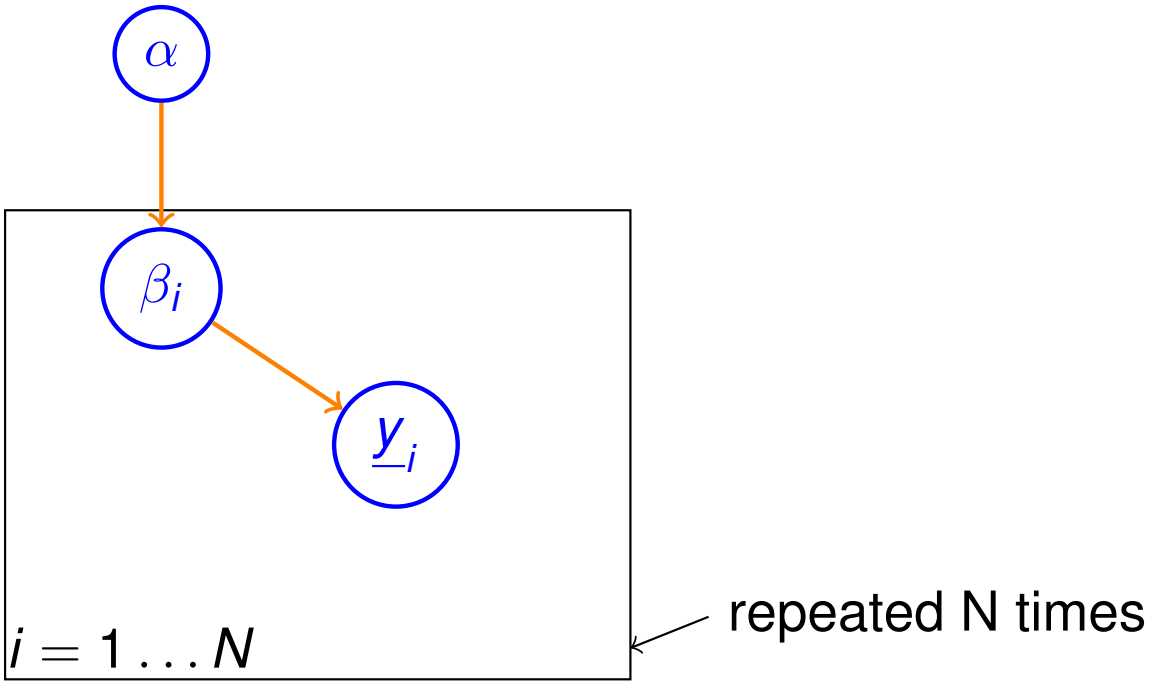


$$P(y|x) = u[0, P(x)] \quad y \sim P(y|x)$$
$$x \sim U[xmin, xmax]$$
$$P(x|y) \propto L(x; y) = \begin{cases} 1 & P(x) \geq y \\ 0 & \text{otherwise} \end{cases}$$

Accept if $L(x; y) = 1$, reject otherwise

Graphically

$$P(\beta_i,\alpha|\{\underline{y}_i\}) \propto \prod_{i=1}^{N} P(\underline{y}_i|\beta_i)P(\beta_i|\alpha)P(\alpha)$$



Example

(Elsner & Jagger 04)

$$y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 CT1 + \beta_2 NAOI + \beta_3 CTI \times NAOI$$

$$\underline{\beta} \sim N(\underline{\mu}, \Sigma^{-1})$$

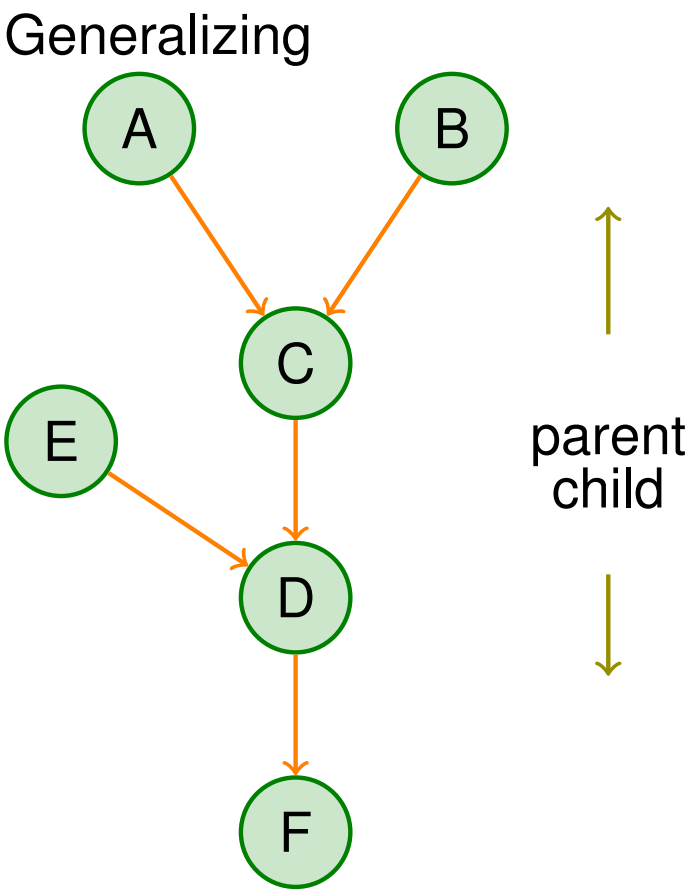
Also read "Regression Machines" for Generalized Linear Models

Constructing priors

A. Conjugate Priors: The Gamma Distribution is a conjugate prior of the Poisson Distribution; so that is one route.
B. Non-informative Prior (flat)
C. Bootstrap-Prior: Use a portion of the data to estimate parameters by MLE.

Other parameter estimates

Frequentist$\Rightarrow$ MLE
(e.g. GLM)



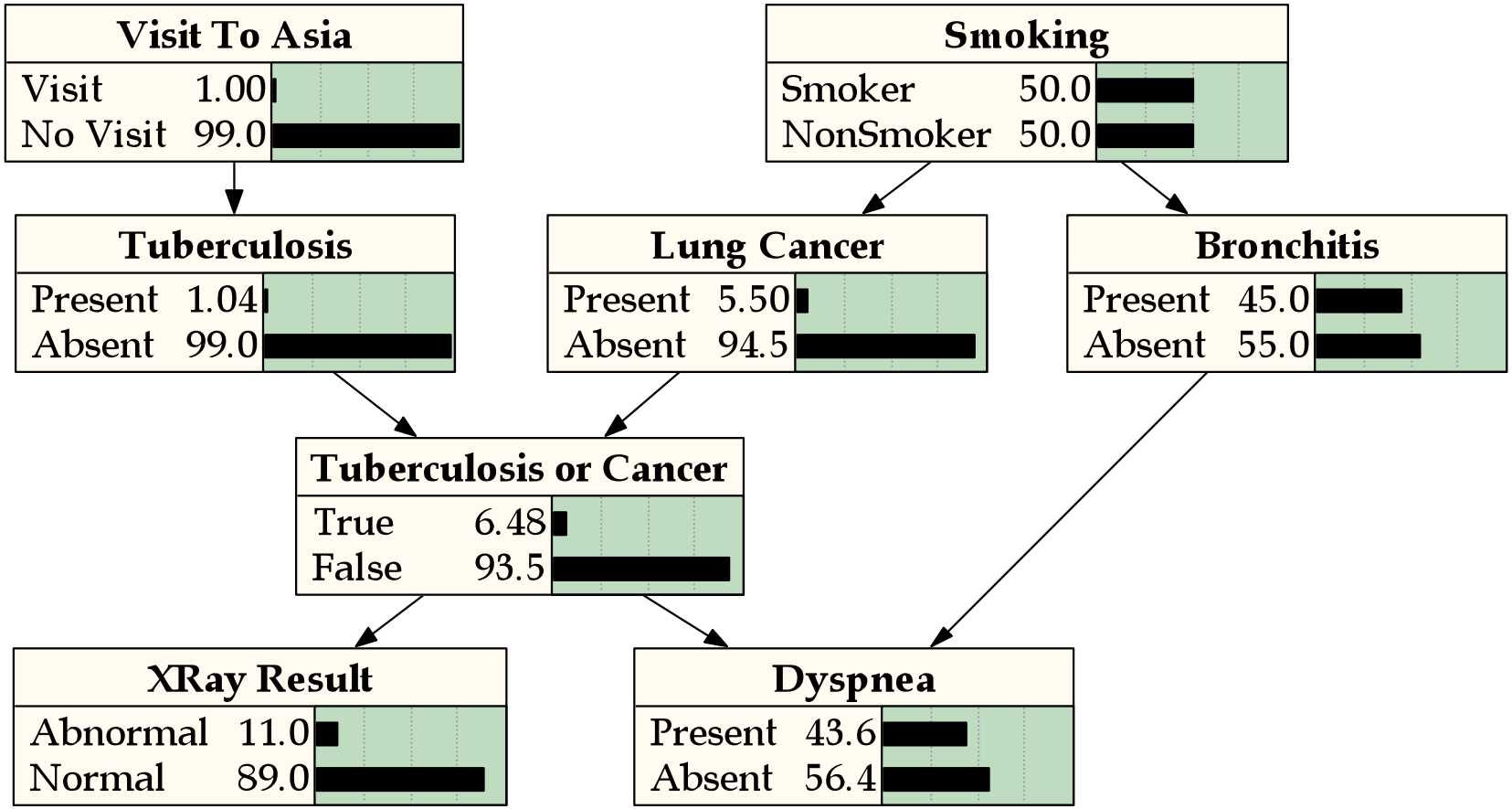
a Hierarchical relationship between variables

b All are random

c Represented by directed acyclic graphs

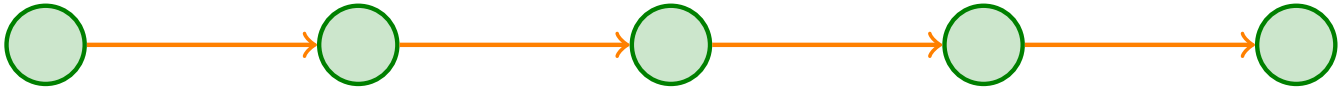
$\Rightarrow$ Bayesian Networks

example

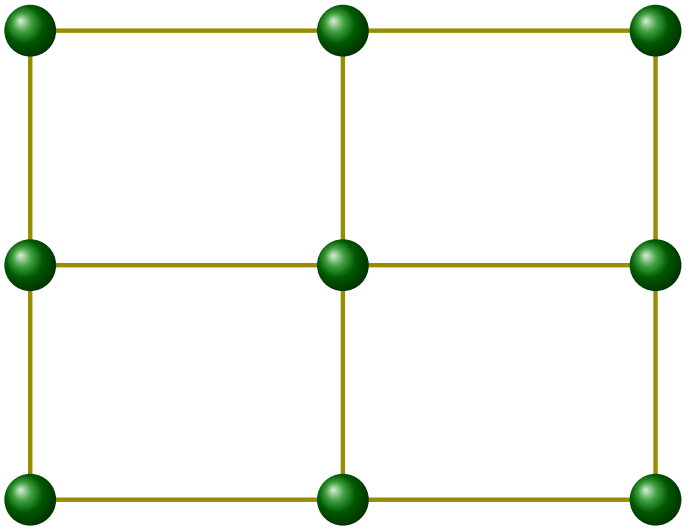

Visit To Asia
Visit 1.00
No Visit 99.0
Smoking
Smoker 50.0
NonSmoker 50.0
Tuberculosis
Present 1.04
Absent 99.0
Lung Cancer
Present 5.50
Absent 94.5
Bronchitis
Present 45.0
Absent 55.0
Tuberculosis or Cancer
True 6.48
False 93.5
XRay Result
Abnormal 11.0
Normal 89.0
Dyspnea
Present 43.6
Absent 56.4

Networking Computer Style

A markov chain is a Bayesian Network



We may model “lattices” through Markov Networks



Markov random field example “ two-way interactions”

Inference on

- Markov Networks
- Bayesian(Belief) Networks

Via

- Graphical Models (see lecture notes).

Daano Ka Filter

- Applied to Sequential filtering problems
- Can also be applied to smoothing problems
- Solution via Recursive Bayesian Estimation
- Approximate Solution
- Can work with non-Gaussian distributions/non-linear dynamics
- Applicable to many other problems e.g. Spatial Inference

Bayesma Pitamah Mantra

$$P(X_k|Y_{1:k}) = \frac{\underbrace{P(Y_k|X_k)}_{2} \underbrace{\sum_{X_{k-1}} P(X_k|X_{k-1})P(X_{k-1}|Y_{1:k-1})}_{1}}{\underbrace{\sum_{X_k}\sum_{X_{k-1}} P(Y_k|X_k)P(X_k|X_{k-1})P(X_{k-1}|Y_{k-1})}_{3}}$$

1. From the Chapman-Kolmogorov equation
2. The measurement model/observation equation
3. Normalization Constant

When can this recursive master equation be solved?

Problem Hai, par usey aur mushkil banao: Mazaa ata hai

How may we relax the Gaussian assumption?

If $P(X_k|X_{k-1})$ and $P(Y_k|X_k)$ are non-gaussian;

How do we represent them, let alone perform these integrations in (2) & (3)?

Daney Filter

Generically

$$P(X) = \sum_{i=1}^{N} w^i \delta(X - X^i)$$

pmf/pdf defined as a weighted sum
$\rightarrow$ Recall from Sampling lecture
$\rightarrow$ Response Surface Modeling lecture

Daney Gin Rahe Hain

In the filtering problem

$$P(X_k|Y_{1:k})$$

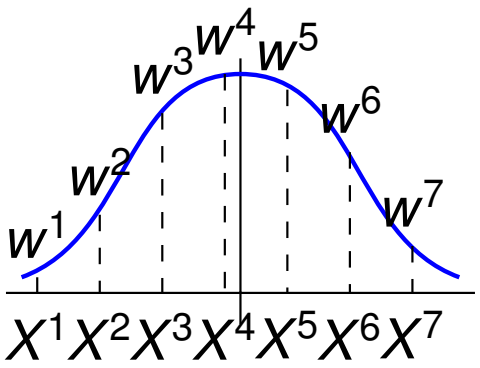
$$w_k^i \propto w_{k-1}^i \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)}$$

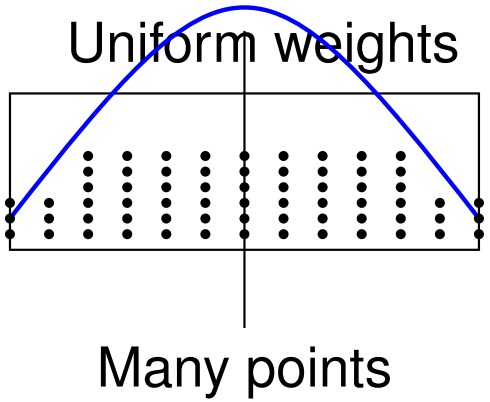
$$(So)\ P(X_k|Y_{1:k}) \cong \sum_{i=1}^{N} w_k^i \delta(X_k - X_k^i)$$

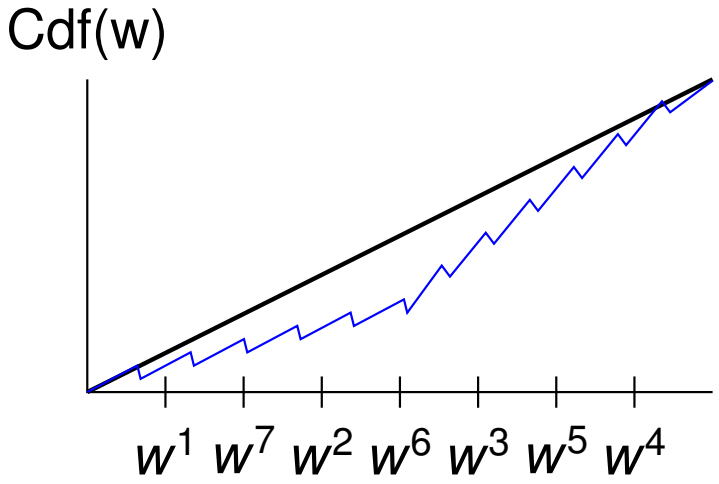
Where the $x_k^i \sim Q(X_k|X_{k-1}^i, Y_k)$
The method essentially draws particles from a proposal distribution and recursively update its weights.

- $\Rightarrow$ No gaussian assumption
- $\Rightarrow$ Neat

Ichuka Dana, Ichuka Dana, Daney ke upar Daana







Resampling ⟶ Sample more probable states more

All Uncertainties killed with the Likelihood Stone

Asymptotically:
$\widehat{Q} \sim P(X_k|X_{k-1}^i) \leftarrow$ Common choice $Q \equiv P(X_k|X_{k-1}^i)$

Sometimes feasible to use proposal from process noise

Then

$w_K^i \propto w_{k-1}^i P(Y_k|X_k^i)$

If resampling is done at every step:

$w_k^i \propto p(Y_k|X_k^i)$

$(w_{k-1}^i \propto \frac{1}{N})$

SIRji, hum bhi hain important

SIR -Sampling Importance Resampling

Input $\{X^i_{k-1}, w^i_{k-1}\}, Y_k$

for $i = 1 : N$

$X^i_k \sim P(X_k|X^i_{k-1})$

$w^i_k = P(Y_k|X^i_k)$

end

$\eta = \sum_i w^i_k$

$w^i_k = w^i_k/\eta$

$\{x^i_k, w^i_k\} \leftarrow$ Resample $[\{X^i_k, w^i_k\}]$

Totally Cooked Up Example

$$X_k = \frac{X_{k-1}}{2} + \frac{25X_{k-1}}{1 + X_{k-1}^2} + 8\cos(1.2k) + v_{k-1}$$

$$Y_k = \frac{X_k^2}{w} + \eta_k$$

$\eta_k \sim N(0, R)$

$v_{k-1} \sim N(0, Q_{k-1})$

Brain Maalish

Sar jo tera chakaraye
Ya matrix dooba jaye
Aja pyare, paas hamare,
Kahe Ghabaraye,
Kahe Ghabaraye
Brain Maalish!

- Covered much, but small portion of the subject
- Computational Atmospheric Statistics at some point in time.
- DO THE PROBLEMS!
- By email, by skype.
- In person in August.

Thanks much for coming, critical feedback welcome!

12.S990 Quantifying Uncertainty

Fall 2012