

統計数理研究所の紹介と 状態空間モデリング

北川源四郎
統計数理研究所

第二回IBM-ISM合同シンポ
2009年5月26日

概略

● 統計数理研究所の紹介

- ・ 統計数理研究所の概要
- ・ 基幹的研究と戦略的研究
- ・ 今後の方向

● 状態空間モデリング

- ・ 状態空間モデル
- ・ 非線形・非ガウス型フィルタ
- ・ 逐次モンテカルロフィルタ

設置目的・沿革

設置目的

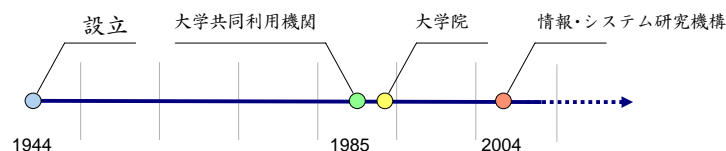
- 統計数理に関する総合研究

沿革

- 1944年：創設
- 1985年：大学共同利用機関に改組転換
- 1988年：総合研究大学院大学創設
- 2004年：法人化、機構化
- 2005年：研究組織再編
- 2009年：立川移転予定

主要活動

- 研究活動
 - ・ 我が国における統計数理の中核拠点
 - ・ 先端的な研究を推進
- 共同研究・共同利用
 - ・ 多様な分野との共同研究
 - ・ スパコン、ソフトウェア、乱数
- 人材育成
 - ・ 総研大等における大学院教育
 - ・ 公開講座

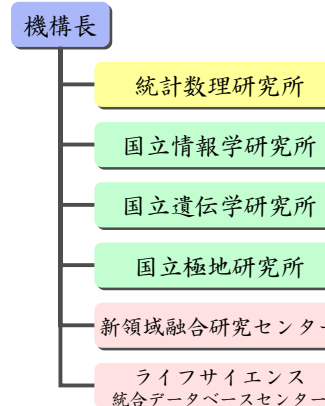


情報・システム研究機構

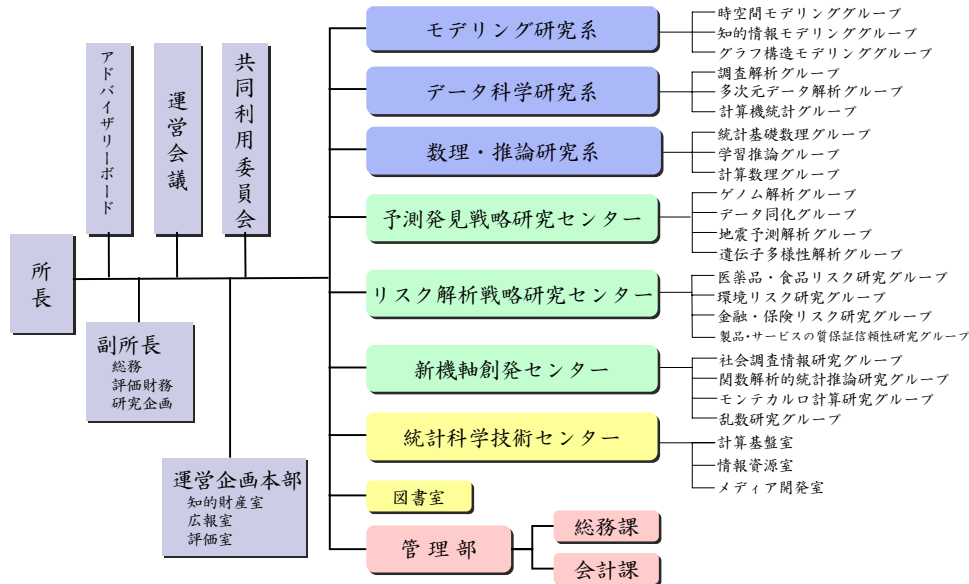
生命、環境、情報社会など、21世紀の人間社会の変容に関わる重要課題の解決には、従来の学問領域の枠にとらわれない研究への取り組みが必要となる。この機構は、**生命、地球、環境、社会など**に関わる複雑な問題を**情報とシステムという立場から捉え**、実験・観測による**大量のデータの生成**とデータベースの構築、**情報の抽出とその活用法**の開発などの課題に関して、分野の枠を越えて融合的に研究すると同時に、新分野の開拓を図ることを目指すものである。

『大学共同利用機関の法人化について』 科学技術・学術審議会 学術分科会、平成15年4月より抜粋

情報・システム研究機構



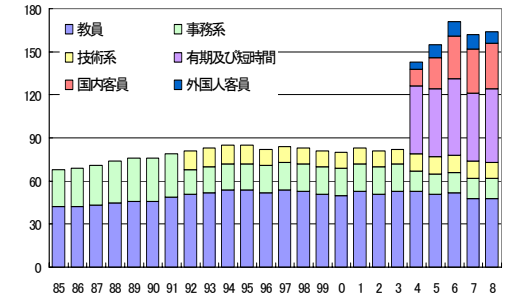
統計数理研究所の組織構成



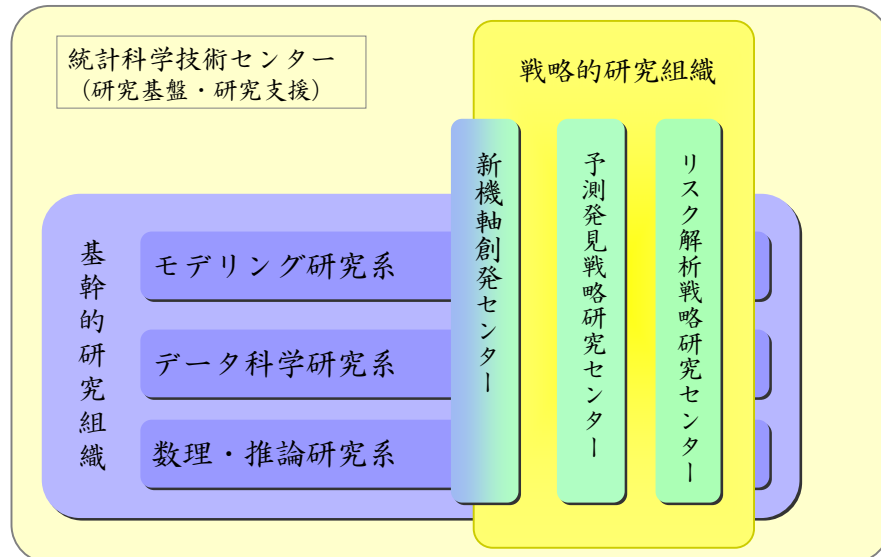
職員数

● 職員数

	H16	H17	H18	H19	H20	H21
教授	19	20	19	16	17	18
准教授	19	18	19	18	16	16
助教	14	12	13	16	15	15
技術職員	12	12	12	13	12	11
管理部	14	14	14	14	14	14
国内客員	10	18	22	29	32	32
外国人客員	5	9	9	10	8	8
有期雇用職員		70	73	54	52	85



研究組織 (基礎研究と戦略的研究)



基幹的研究組織

モデリング研究系

多数の要因が複雑に関連した、時間的・空間的に変動する現象や知的な情報処理のモデル化、およびモデルに基づく統計的推論の方法に関する研究

時空間モデリンググループ
 知的情報処理モデリンググループ
 グラフ構造モデリンググループ

データ科学研究系

不確実性と情報の不完全性に対処するためのデータ設計と調査、分析の方法および計算機の高度利用に基づくデータ解析法に関する研究

調査解析グループ
 多次元データ解析グループ
 計算機統計グループ

数理・推論研究系

統計基礎理論および統計的学習理論の構築、および統計的推論に必要な最適化、計算アルゴリズムの理論と基礎数理に関する研究

統計基礎数理グループ
 学習推論グループ
 計算数理グループ

Current Emphases in Strategic Research

1. Information Society

IT + Information oriented society
Extraction of information from massive large scale data

➡ Prediction and Knowledge Discovery

2. Society with Risk & Uncertainty

Increase of uncertainty and risks due to globalization of society
(finance, economy, pension, environment, drug and food safety disaster prevention,)

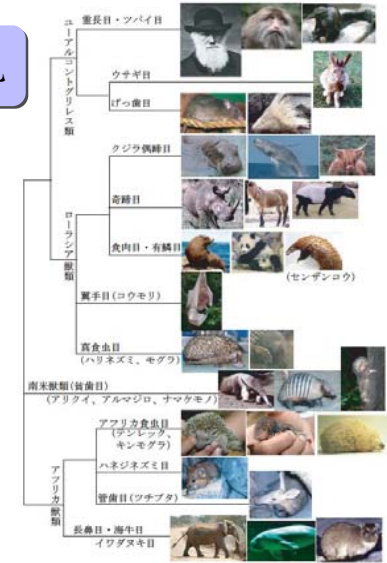
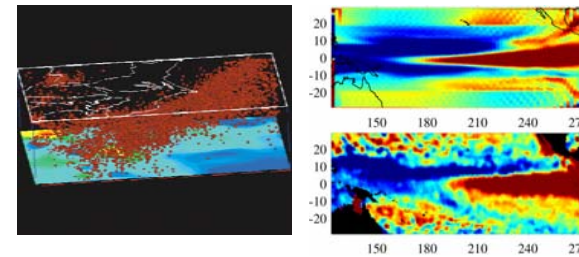
Scientific approach to risk management

➡ Modeling of uncertainty & managing risks

予測発見戦略研究センター

大規模データに基づく予測と知識発見

データ同化グループ
地震予測解析グループ
ゲノム解析グループ
生物多様性解析グループ



リスク解析戦略研究センター

背景

社会のグローバル化によって不確実性とリスクが増大

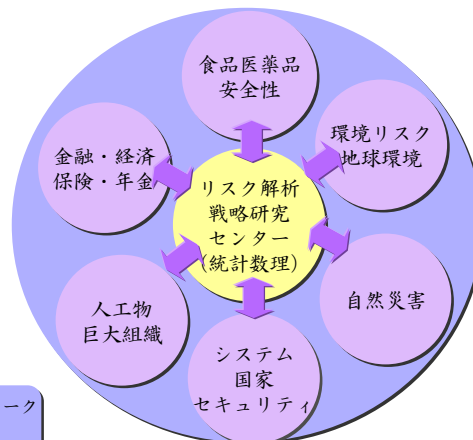
目的

リスクの定量的評価法の確立 ➡ リスクの科学的管理

活動内容

- (1) リスク評価・管理の方法の研究拠点整備
- (2) プロジェクト研究推進
 - 医薬品・食品リスク
 - 環境リスク
 - 金融・保険リスク
 - 製品・サービスの質保証・信頼性
- (3) リスク研究のNOE(Network of Excellence)形成
- (4) 研究現場でのT型人材育成

Network of Excellence



リスク研究ネットワーク

(39組織)	
研究科	6
研究所・センター	11
学協会	12
その他団体	10

統計数理の汎用性・学際性を活かし、
リスク研究における要の役割を果たす

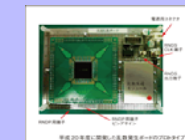
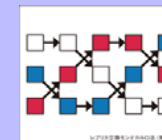
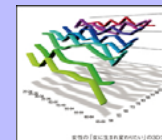
新機軸創発センター

設置目的

社会および学術研究の新しい潮流を見据え、萌芽的なものも含めて自由な発想に基づく研究プロジェクトを推進することにより、統計数理における新機軸の確立を目指す。

研究グループ

- ・社会調査情報研究グループ
- ・関数解析的統計推論研究グループ
- ・モンテカルロ計算研究グループ
- ・乱数研究グループ



Function and Induction Research Project

A Research Project at Transdisciplinary Research Integration Center, ROIS

1. Prediction and Risk Analysis

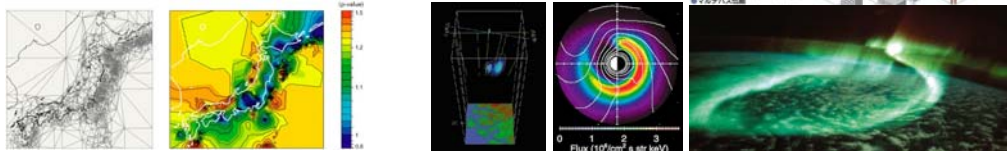
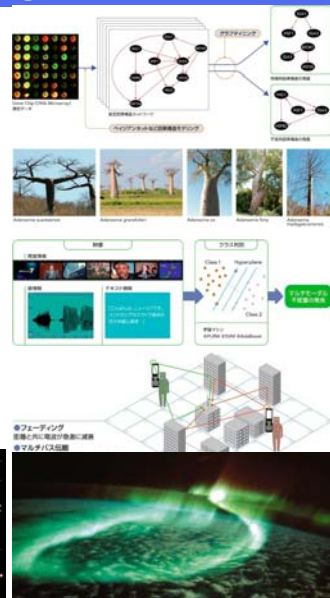
- 1-1 Discovery of causal networks by graph mining
- 1-2 Statistical analysis of genome diversity
- 1-3 Toward holistic understanding of biological diversity
- 1-4 Inductive inference and functional design of statistical analysis for risk evaluation

2. Metaware for Information and Communication

- 2-1 Computer environment for inductive modeling
- 2-2 Discovery of invariants in multimodal data
- 2-3 Wireless communication system based on modeling and high-performance data processing

3. Dynamic Inverse Problems

- 3-1 Inverse analysis method for earth science based on statistical modeling
- 3-2 Inductive analysis of multidimensional array data for understanding complex systems and predicting their future
- 3-3 Research on seismic activity by statistical model



The Institute of Statistical Mathematics

13

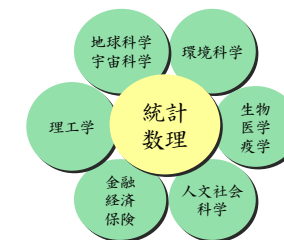
公募型共同研究

共同研究採択件数・研究者数

年度	件数	研究者数
2003	99	565
2004	108	595
2005	124	652
2006	122	626
2007	120	663
2008	124	553

多様な分野との共同研究

- ・ 数理的方法, モデリング
 - ・ データ解析法
 - ・ 豊富な人的ネットワーク
 - ・ 知識交流のハブ
 - ・ 充実した計算機設備
-
- ・ 統計数理研究者との共同研究
 - ・ 広範な領域の研究者との共同研究
理工学, 宇宙・地球科学, 生物・医学,
人文・社会科学, 環境科学
 - ・ 萌芽的な研究の重視



The Institute of Statistical Mathematics

14



総合研究大学院大学における教育

総合研究大学院大学とは

- 昭和63年, 全国の大学共同利用機関を基盤機関として創設。
当初は2研究科8専攻でスタート
- 平成16年, 法人化を機に現在の6研究科22専攻体制に再編。本部: 神奈川県葉山
文化科学研究科
物理科学研究科
高エネルギー加速器科学研究科
複合科学研究科
生命科学研究所
先端科学研究科
- 総合研究大学院大学の特色
 - ・ 全国初の独立大学院大学
 - ・ 博士後期課程と5年一貫制の併設
 - ・ 18大学共同利用機関等との連携・協力

複合科学研究科

- 情報・システム研究機構の研究所が基盤機関
 - ・ 統計科学専攻 (統計研)
 - ・ 情報学専攻 (情報研)
 - ・ 極域科学専攻 (極地研)
- 平成18年度より5年一貫制博士課程を開始

統計科学専攻

- 大学創設時から参加
- 我が国唯一の統計科学の専攻
- 入学定員5名
(5年一貫制2名, 後期3名)
在籍数25名
- 学位授与者数: 平成19年までに77名

The Institute of Statistical Mathematics

15

立川市への移転 (平成21年9月)

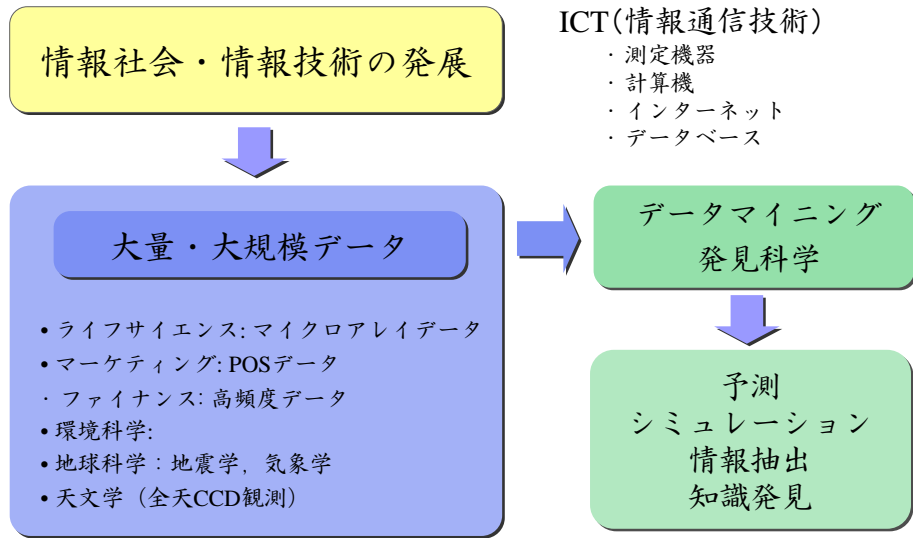
移転場所: 立川市緑町
土地面積: 62,450 m²
建設面積: 57,550 m²
総合研究棟 I, II
交流棟 I, II (宿泊, 展示)
移転予定: H21年9月



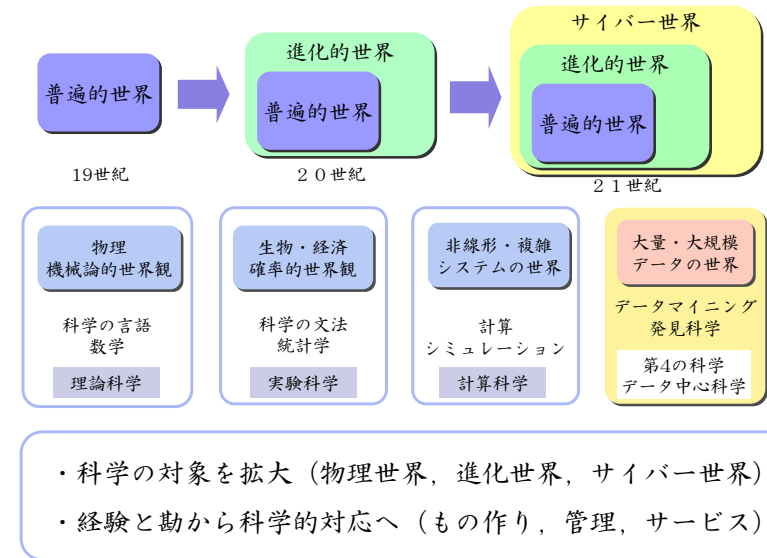
The Institute of Statistical Mathematics

16

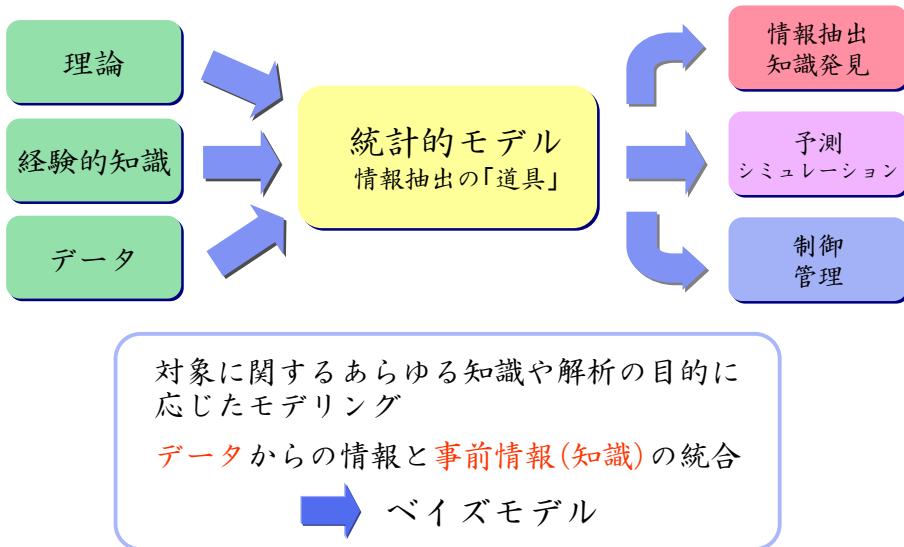
Post-IT時代と大量・大規模データ



科学の対象と方法論の拡大



能動的なモデリング



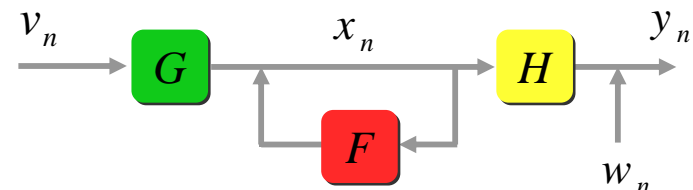
状態空間モデル

$$x_n = Fx_{n-1} + Gv_n \quad \text{状態モデル}$$

$$y_n = Hx_n + w_n \quad \text{観測モデル}$$

y_n 時系列 v_n システムノイズ

x_n 状態ベクトル w_n 観測ノイズ

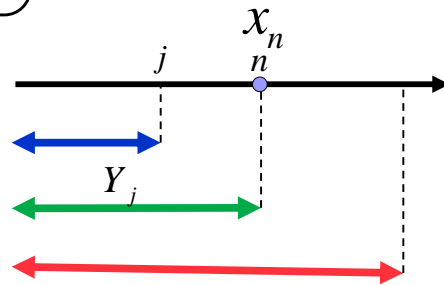


State Estimation Problem

Conditional distribution
of x_n given Y_j

Observations
 $Y_j \equiv \{y_1, \dots, y_j\}$

- for $j < n$ Prediction
- for $j = n$ Filter
- for $j > n$ Smoother



Kalman Filter

予測

$$x_{n|n-1} = F_n x_{n-1|n-1}$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T$$

フィルタ

$$K_n = V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + R_n)^{-1}$$

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1})$$

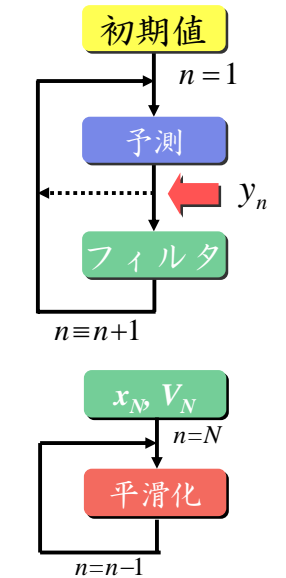
$$V_{n|n} = (I - K_n H_n) V_{n|n-1}$$

平滑化

$$A_n = V_{n|n} F_n^T V_{n+1|n}^{-1}$$

$$x_{n|N} = x_{n|n} + A_n (x_{n+1|N} - x_{n+1|n})$$

$$V_{n|N} = V_{n|n} + A_n (V_{n+1|N} - V_{n+1|n}) A_n^T$$



状態空間モデルの応用

● 定常時系列モデルの統一的取り扱い

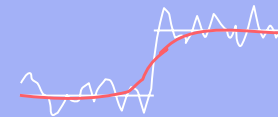
- ・ 時系列モデルの最尤推定 (ARMAモデルなど)
- ・ 欠測値の補間、予測分布
- ・ 制御への応用

● 非定常性のモデリング

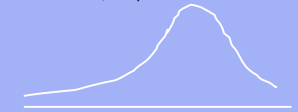
- ・ トレンド推定
- ・ 季節調整
- ・ 確率的ボラティリティ
- ・ 時変係数ARモデル (非定常スペクトル)
- ・ 信号抽出

非ガウス型モデリングの必要性

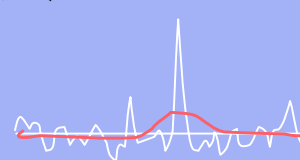
構造変化



非対称分布



異常値 (外れ値)



非線形性

$$x_n = f(x_{n-1}) + v_n$$

離散過程

ポアソン過程
二項過程

状態空間モデルの拡張

線形・ガウス型

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= Hx_n + w_n \end{aligned}$$

非線形・非ガウス型

$$\begin{aligned} x_n &= f(x_{n-1}, v_n) \\ y_n &= h(x_n, w_n) \end{aligned}$$

一般型

$$\begin{aligned} x_n &\sim F(\cdot | x_{n-1}) \\ y_n &\sim H(\cdot | x_n) \end{aligned}$$

関数：非線形
分布：非ガウス型

条件付分布
離散状態・離散観測値

非ガウス型フィルタ・平滑化

一期先予測

$$p(x_n | Y_{n-1}) = \int_{-\infty}^{\infty} p(x_n | x_{n-1}) p(x_{n-1} | Y_{n-1}) dx_{n-1}$$

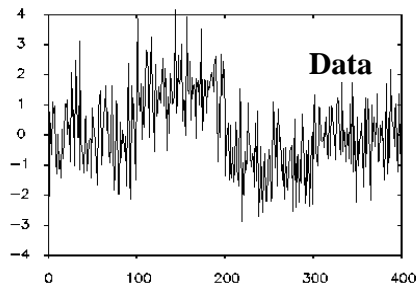
フィルタ

$$p(x_n | Y_n) = \frac{p(y_n | x_n) p(x_n | Y_{n-1})}{p(y_n | Y_{n-1})}$$

平滑化

$$p(x_n | Y_N) = p(x_n | Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1} | x_n) p(x_{n+1} | Y_N)}{p(x_{n+1} | Y_n)} dx_{n+1}$$

レベルシフトの自動検出

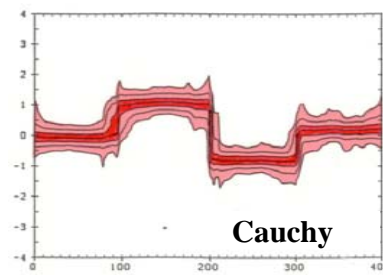
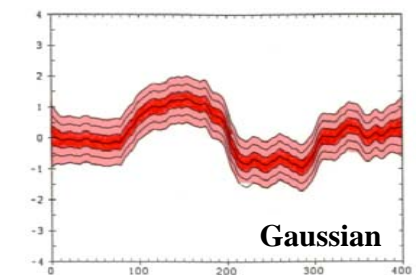


トレンドモデル

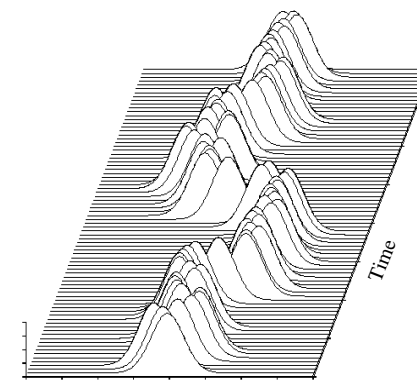
$$\begin{aligned} t_n &= t_{n-1} + v_n \\ y_n &= t_n + w_n \end{aligned}$$

ノイズ分布

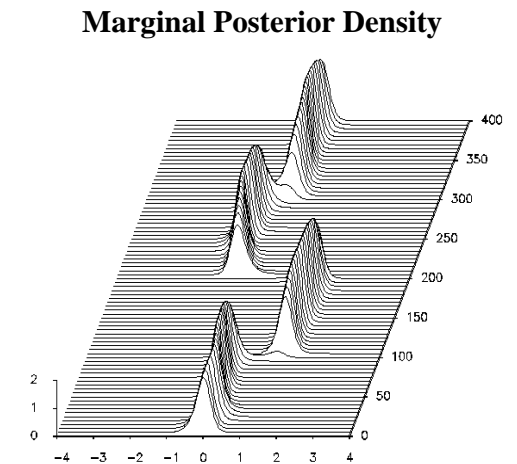
$$\begin{aligned} v_n &\sim C(0, \tau^2) \quad \text{コーシー分布} \\ w_n &\sim N(0, \sigma^2) \quad \text{正規分布} \end{aligned}$$



非ガウス型フィルタ・平滑化



Gaussian Distribution



Cauchy Distribution

自己組織型フィルタ・平滑化

状態空間モデル

$$x_n = x_{n-1} + v_n$$

$$y_n = x_n + w_n$$

コーシー分布

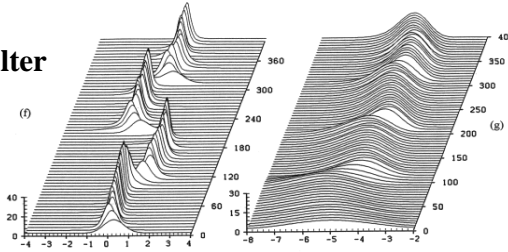
$$p(v_n) = \frac{\tau}{\pi} \frac{1}{(v_n^2 + \tau^2)}$$

Augmented State Space Model

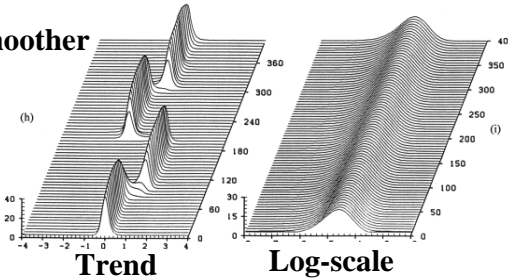
$$\begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \log \tau_{n-1}^2 \end{bmatrix} + \begin{bmatrix} \tau_{n-1} \\ 0 \end{bmatrix} v_n$$

$$y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} + w_n$$

Filter



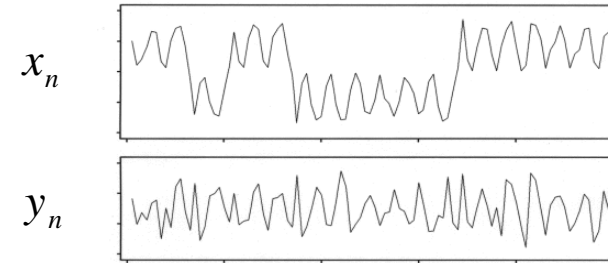
Smoother



非線形モデル

$$x_n = \frac{1}{2} x_{n-1} + \frac{25 x_{n-1}}{1 + x_{n-1}^2} + 8 \cos(1.2n) + v_n$$

$$y_n = \frac{x_n^2}{20} + w_n \quad v_n \sim N(0,0.1), w_n \sim N(0,1)$$

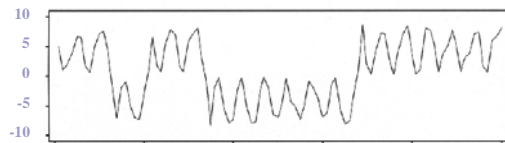


Signal

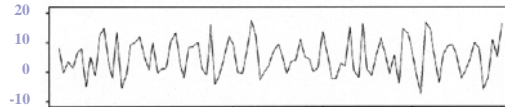
Observation

非線形フィルタ・平滑化

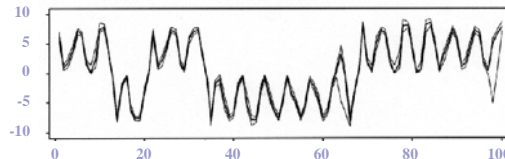
Signal



Observed



Estimated



分布の近似

0. 正規分布近似

(拡張) カルマンフィルタ・平滑化

1. 区分線形 (階段) 近似

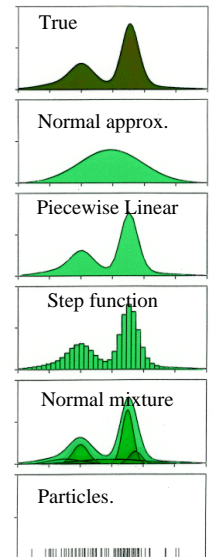
非ガウス型フィルタ・平滑化

2. 混合正規分布近似

ガウス和フィルタ・平滑化

3. 粒子近似

逐次モンテカルロフィルタ・平滑化

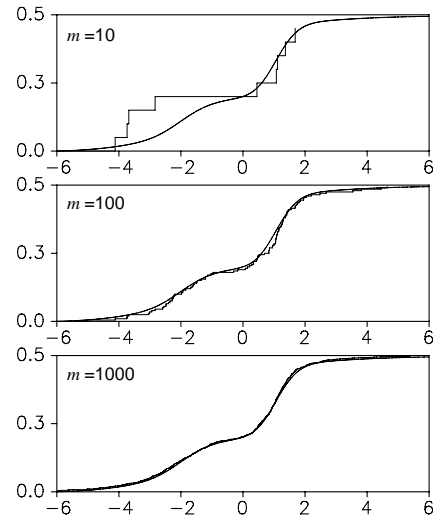
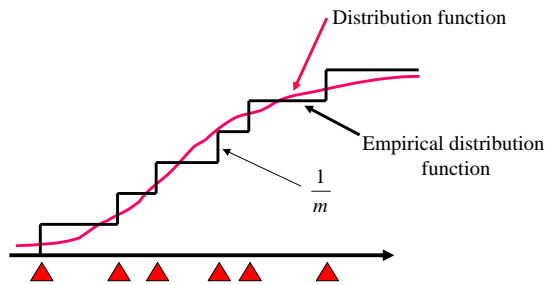


分布関数の粒子による近似

$$\{p_n^{(1)}, \dots, p_n^{(m)}\} \sim p(x_n | Y_{n-1})$$

$$\longleftrightarrow F_n(x) = \frac{1}{m} \sum_{j=1}^m I(x; p_n^{(j)})$$

$$\longleftrightarrow \Pr(X_n = p_n^{(j)} | Y_{n-1}) = \frac{1}{m}$$



分布の粒子による近似

$$\{p_n^{(1)}, \dots, p_n^{(m)}\} \sim p(x_n | Y_{n-1}) \quad \text{Predictive Dist.}$$

$$\{f_n^{(1)}, \dots, f_n^{(m)}\} \sim p(x_n | Y_n) \quad \text{Filter Dist.}$$

$$\{s_{n|N}^{(1)}, \dots, s_{n|N}^{(m)}\} \sim p(x_n | Y_N) \quad \text{Smoother Dist.}$$

$$\{v_n^{(1)}, \dots, v_n^{(m)}\} \sim p(v_n) \quad \text{System Noise Dist.}$$

Sequential Monte Carlo Methods

Also called "Particle Filter"

Gordon, Salmond and Smith (1993)

Bootstrap Filter

Kitagawa (1993, 1996)

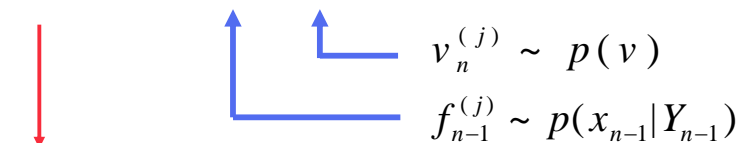
Monte Carlo Filter/Smoother

Doucet, de Freitas and Gordon (2001)

"Sequential Monte Carlo Methods in Practice"

一期先予測

$$x_n = F(x_{n-1}, v_n) \quad \text{システムモデル}$$



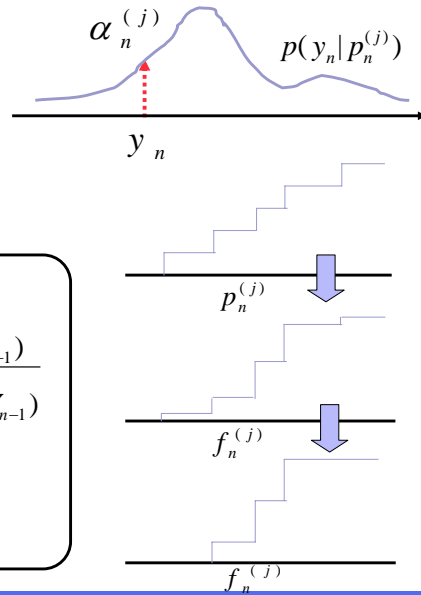
$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$$

フィルタ (リサンプリング)

$\alpha_n^{(j)}$: 粒子 $p_n^{(j)}$ のベイズ係数

$$\alpha_n^{(j)} = p(y_n | X_n = p_n^{(j)})$$

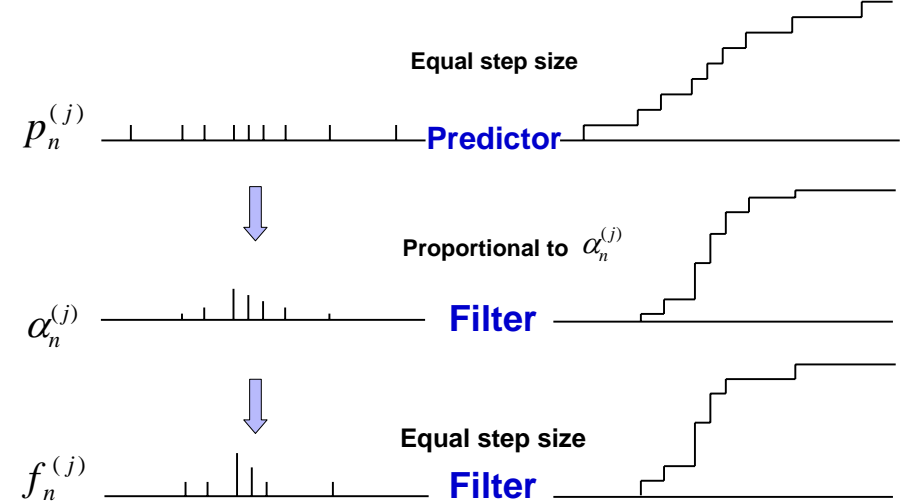
$$\begin{aligned} \Pr(X_n = p_n^{(j)} | Y_n) &= \Pr(X_n = p_n^{(j)} | Y_{n-1}, y_n) \\ &= \frac{\Pr(y_n | X_n = p_n^{(j)}) \Pr(X_n = p_n^{(j)} | Y_{n-1})}{\sum_{i=1}^m \Pr(y_n | X_n = p_n^{(i)}) \Pr(X_n = p_n^{(i)} | Y_{n-1})} \\ &= \frac{\alpha_n^{(j)} \frac{1}{m}}{\sum_{i=1}^m \alpha_n^{(i)} \frac{1}{m}} = \frac{\alpha_n^{(j)}}{\sum_{i=1}^m \alpha_n^{(i)}} \end{aligned}$$



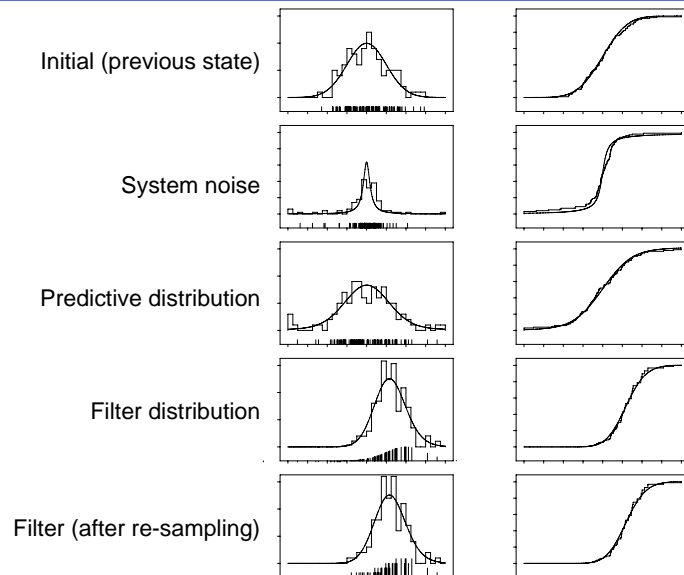
Filtering

Weights of Particles

Empirical Distribution



MCF の 1 サイクル



逐次モンテカルロ・フィルタ

- ・ システムノイズ

$$v_n^{(j)} \sim p(v) \quad j = 1, \dots, m$$

- ・ 予測分布

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$$

- ・ 重要度 (ベイズ係数)

$$\alpha_n^{(j)} = p(y_n | p_n^{(j)})$$

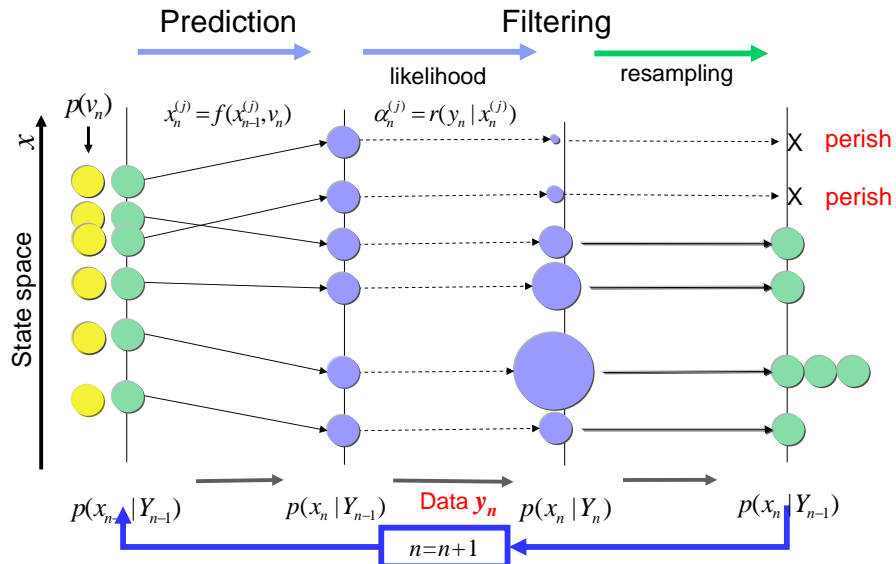
- ・ フィルタ分布のリサンプリング

$$\{p_n^{(j)}\} \rightarrow \{f_n^{(j)}\}$$

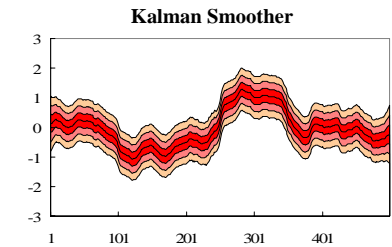
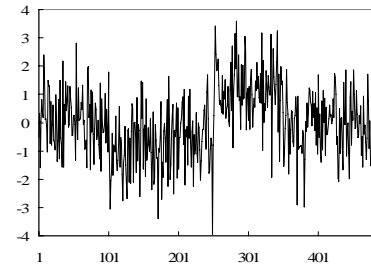
Gordon et al. (1993), Kitagawa (1996)

Doucet, de Freitas and Gordon (2001) "Sequential Monte Carlo Methods in Practice"

One Cycle of Monte Carlo Filtering



非ガウス型平滑化



Trend Model

$$t_n = t_{n-1} + v_n$$

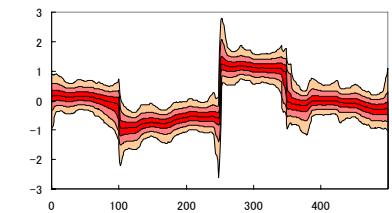
$$y_n = t_n + w_n$$

Noise Distribution

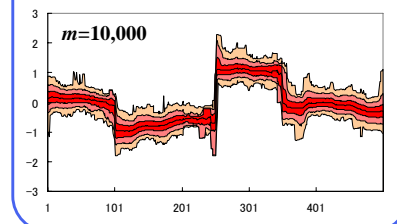
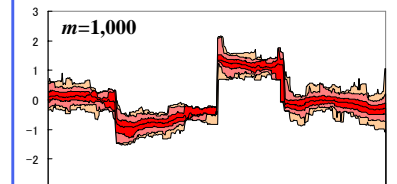
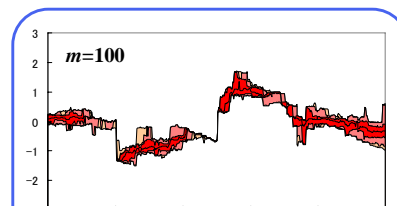
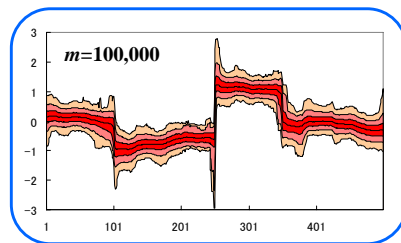
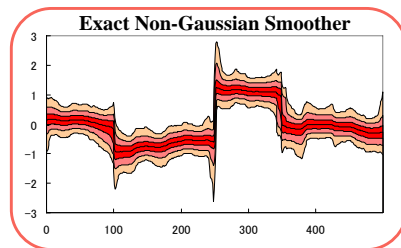
$$v_n \sim N(0, \tau^2) \text{ or } C(0, \tau^2)$$

$$w_n \sim N(0, \sigma^2)$$

Exact Non-Gaussian Smoother



Single MCF



Applications of MCF

1. **Non-Gaussian smoothing**
 - Level shift
 - Non-Gaussian seasonal adjustment
 - Stochastic volatility models
2. **Nonlinear smoothing**
 - Tracking
 - Phase-unwrapping
3. **Signal extraction problems**
4. **Modeling count data**
5. **Self-organizing state space model**
6. **High-dimensional filtering/smoothing**

Gordon et al. (1993), Kitagawa (1996)

Doucet, de Freitas and Gordon (2001) "Sequential Monte Carlo Methods in Practice"