知的情報処理論 第11回

2023年6月27日（火）
武田

お断り

用語
- おおよそ連続確率変数が前提
  確率，分布，確率密度 → 確率密度関数の意
  e.g. 事前分布 = 事前確率 = 事前確率密度

表記: 断りがない場合
- 大文字 bold → 適当な次元の行列 $\mathbf{A}$
- 小文字 bold → 適当な次元の（列）ベクトル $\mathbf{a}$
- (離散的な)時刻 a から b をまとめた行列
  （ベクトルの系列）$\mathbf{a}_{a:b} = [\mathbf{a}_a, \mathbf{a}_{a+1}, \ldots, \mathbf{a}_b]$

各回の内容（予定）

第9回
導入

適応フィルタ

第10回
エコーキャンセラ
LMS
NLMS
RLS

第11回
トラッキング
Kalman Filter
Particle Filter
・**式ばかりで**
一番つらい回
・**話の抽象度も高い**
☹☹☹

第12回
音源分離
ICA　PCA
IVA　T-F Mask

音声認識

第13回
概要 + 音響モデル
GMM-HMM
DNN-HMM

第14回
音響モデル + 言語モデル
N-gram
RNN

第15回
言語モデル + 探索
WFST

各回の内容（予定）

第9回
導入

一般化

第10回
再帰的な更新
（逐次処理）

第11回
状態空間モデル
・連続変数での
状態推定
（**ベイズ推定の観点**）

ここより右は
学習を含む

第12回
・連続変数の状態推定
・時刻間での独立性
・パラメータ学習

音声認識

第13回
状態空間モデル
・離散変数での状態推定
・パラメータ学習

第14回
・離散変数
・パラメータ学習

第15回
リアルタイム処理

頭の片隅に置いとく方が良い大まかな要素

コスト関数
- 「良さ」を測る関数
  教師あり学習: 二乗誤差, CrossEntropy, etc…
  教師なし学習: 尤度, 再構成誤差，etc…
  認識・推定: 期待値, 最大事後確率(MAP), etc…
- ロス/目的関数ともいう

モデル
- 入出力の関係や拘束条件を記述
- 方程式や確率モデルで表現
  （コスト関数と一体化 or 切り離せない場合も）

ざっくりとした
認識/学習/探索
- コスト関数を{最大化|最小化|良く}する値（集合）を求める手続き（**最適化等の分野**）
- モデル構造や補助変数・関数を利用した手続きも存在

「局所的最適解（局所解）」「大域的最適解」

予備: 誤差と対数尤度

例: 回帰モデル

線形　$\hat{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x}$　$\mathbf{w}$ 重み　$\mathbf{x}$ 既知データ　（非線形でも同様　$\hat{y} = \boldsymbol{f}(\mathbf{x}; \boldsymbol{\Theta})$）

下記2つのコスト関数はほぼ等価
- 二乗誤差コスト　$(y - \mathbf{w}^{\mathrm{T}}\mathbf{x})^2$　　y 観測データ/正解（固定値）
- ガウス雑音(or **誤差**) + 負の対数尤度

$$y = \underbrace{\mathbf{w}^{\mathrm{T}}\mathbf{x}}_{\text{固定値}} + \underbrace{\varepsilon}_{\text{ばらつく値（誤差）}} \qquad \varepsilon \sim N(0, \sigma^2)$$

⇩

負の対数尤度　$$-\log p(y \mid \mathbf{x}, \mathbf{w}, \sigma^2) = \frac{1}{2}\log 2\pi\sigma^2 + \frac{1}{2\sigma^2}\underbrace{(y - \mathbf{w}^{\mathrm{T}}\mathbf{x})^2}_{\text{二乗誤差部分}}$$

他: L1ノルム と ラプラス分布

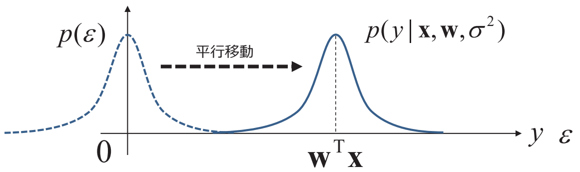
予備：誤差と対数尤度

不確定要素込みで上手く表現したい
⇩
"どの値を取りやすいか" 分布で表現

$$y = \underbrace{\mathbf{w}^{T}\mathbf{x}}_{\text{固定値＝定数扱い}} + \underbrace{\varepsilon}_{\text{不確定要素}} \qquad \varepsilon \sim N(0, \sigma^2)$$

$$p(y \mid \underbrace{\mathbf{x}, \mathbf{w}, \sigma^2}_{\text{条件部分：この3つを仮定すれば，y の分布が定義できる}}) = N(y \mid \mathbf{w}^{T}\mathbf{x}, \sigma^2)$$

ε の分布から平均がズレた分布＝ $p(y \mid \mathbf{x}, \mathbf{w}, \sigma^2)$

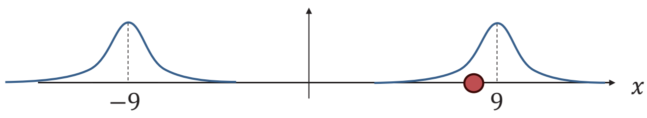


ε（雑音/誤差）がない場合，平たくいうと "デルタ関数" とも捉えられる $\delta(y - \mathbf{w}^{T}\mathbf{x})$

予備：尤度

2つのガウス分布（２つのクラスタ）

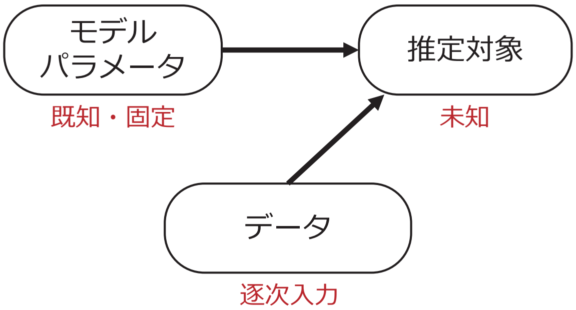
$p(x|z=-9) = \mathcal{N}(x; -9, 1)$ $\qquad$ $p(x|z=9) = \mathcal{N}(x; 9, 1)$



例： データ $x = 8$ はどちらから（$z = 9$ or $z = -9$ の設定）生成されたと考えるのが **<u>尤もらしい</u>** か

– x は両方で共通かつ**固定の値（ここでは 8）**

今回の話題



本日の内容：適応フィルタ

再帰的なベイズ推定で問題を捉える

前回：エコーキャンセラ問題，RLS法

↓ 動的な状況，一般化，再帰的推定

1. <u>一般的な動的システムのモデル</u> **概念やイメージを掴むのが目標**

↓ 線形・ガウスの場合

やややマク

2. Kalman Filter

モンテカルロ法の場合

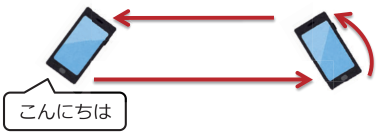
3. Particle Filter

非ガウス，
離散変数の場合

音源分離，音声認識など

エコーキャンセラ（適応フィルタ）

自分の声（既知）が回り込む



音声が回り込む
⇩
明瞭度の低下
エコーの増幅（ハウリング）

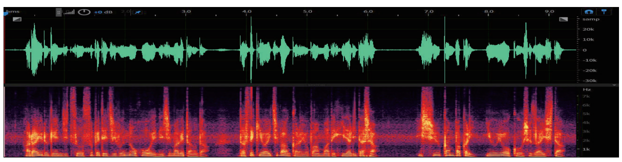
自分の声（既知）が回り込む

デバイスの声も入力
⇩
音声認識が困難

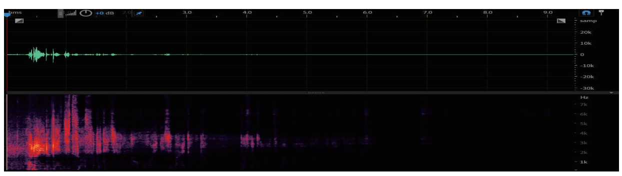
エコー（既知音）を抑圧したい

エコーキャンセラ

キャンセル前（観測信号）

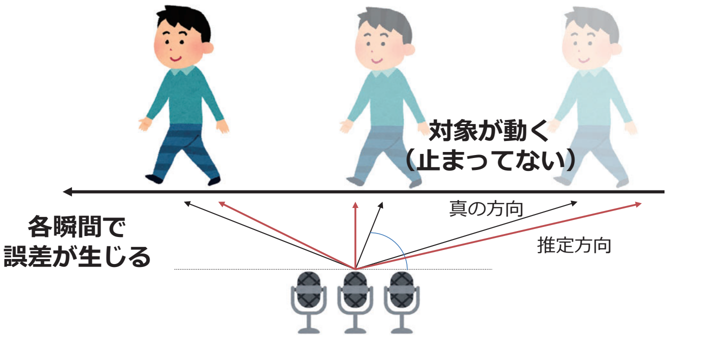


逐次的なキャンセル



トラッキング
斜方投射 + 測定雑音

移動音源のトラッキング



誤差少なく音源方向（位置）を追従したい

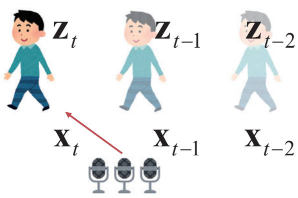
音源定位
(各瞬間瞬間での方向推定)

具体例 → 一般モデル → 事後分布による再帰推定

一般的な動的システムのモデル

具体的な例: トラッキング

目的: 移動する音源追跡
入力 $\mathbf{x}_t$: センサデータ/推定音源位置
推定対象 $\mathbf{z}_t$: 音源座標・速度・加速度など
前提: 対象がどう動くのかは少しは予測できる



具体的な例: トラッキング

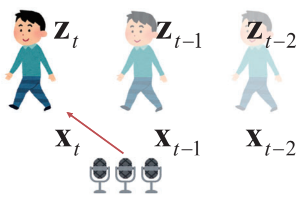
目的: 移動する音源追跡
入力 $\mathbf{x}_t$: センサデータ/推定音源位置
推定対象 $\mathbf{z}_t$: 音源座標・速度・加速度など
前提: 対象がどう動くのかは少しは予測できる

モデル化

例えば Δt 時間間隔

座標	x_t	$= x_{t-1} + v_{t-1}\Delta t$
速度	v_t	$= v_{t-1} + a_{t-1}\Delta t$
加速度	a_t	$= a_{t-1} + w_{t-1}$
	$\mathbf{z}_t$	$\mathbf{z}_{t-1}$

不確定要素
/駆動雑音



具体的な例: トラッキング

目的: 移動する音源追跡
入力 $\mathbf{x}_t$: センサデータ/推定音源位置
推定対象 $\mathbf{z}_t$: 音源座標・速度・加速度など
前提: 対象がどう動くのかは少しは予測できる

モデル化

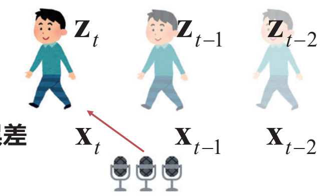
例えば

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{w}_{t-1}$$

$$\mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{v}_t$$

（こう表せるとする）

雑音や
モデル化誤差



具体的な例: トラッキング

目的: 移動する音源追跡
入力 $\mathbf{x}_t$: センサデータ/推定音源位置
推定対象 $\mathbf{z}_t$: 音源座標・速度・加速度など
前提: 対象がどう動くのかは少しは予測できる

例えば

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{w}_{t-1}$$

$$\mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{v}_t$$

・対象がどう変化するか
・データがどう観測されるか

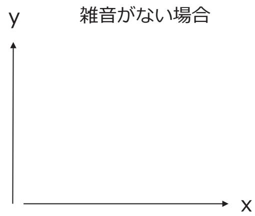
これらのモデルを仮定して，直接観測できない $\mathbf{z}_t$ をデータ $\mathbf{x}_t$ から動的に推定

仮想的な例

2次元斜方投射

– 適当な初速を与えて，斜めにボールを投げる
– 重力加速度の影響で軌跡は放物線を描く

ただし，観測データにはノイズが重畳



仮想的な例

2次元斜方投射 (y軸成分)

推定対象 $\mathbf{z}_t$: 座標 c_t，速度 v_t ，加速度 a_t
モデル: 運動方程式（大雑把に離散時間近似）

y軸方向の状態遷移モデル

Δt 時間間隔

$$c_t = c_{t-1} + v_{t-1}\Delta t + \boxed{w_{c,t-1}}$$
$$v_t = v_{t-1} + a_{t-1}\Delta t + \boxed{w_{v,t-1}}$$
$$a_t = -g$$

誤差/
駆動雑音項

重力加速度

※ 実験データから求める＝学習

$$\underbrace{\begin{bmatrix} c_t \\ v_t \\ a_t \end{bmatrix}}_{\mathbf{z}_t} = \underbrace{\begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} c_{t-1} \\ v_{t-1} \\ -g \end{bmatrix}}_{\mathbf{z}_{t-1}} + \underbrace{\begin{bmatrix} w_{c,t-1} \\ w_{v,t-1} \\ 0 \end{bmatrix}}_{\mathbf{w}_{t-1}}$$

仮想的な例

2次元斜方投射 (y軸成分)

推定対象 $\mathbf{z}_t$: 座標 c_t，速度 v_t ，加速度 a_t
モデル: 運動方程式（大雑把に離散時間近似）

y軸方向の観測モデル

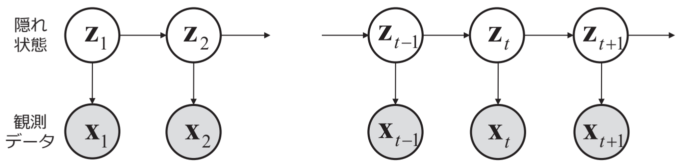
$$y_t = c_t + v_t$$

雑音 v_t

$$\underbrace{y_t}_{\mathbf{x}_t} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} c_{t-1} \\ v_{t-1} \\ -g \end{bmatrix}}_{\mathbf{z}_t} + \underbrace{v_t}_{\mathbf{v}_t}$$

一般的な動的システムのモデル

状態空間モデル: state space model



- **「状態」と「観測」の2つの概念**
- **状態に基づいてデータが生成される過程をモデル化（ランダム要素を含む．確率的生成モデルの一つ）**
- **予測の信頼度も計算可能**

一般的な動的システムのモデル

状態空間モデル: state space model

– 状態ベクトル $\mathbf{z}_t$
  - (あらゆる)知りたい情報を表すベクトル
    例: 座標・速度，文字列，クラス，全context，etc…
  - 直接は観測できない: **潜在変数**

– プロセス/状態方程式 $\mathbf{z}_t = \mathbf{f}_{t-1}(\mathbf{z}_{t-1}, \mathbf{w}_{t-1})$
  - 状態が**どう変化するか**を表現
  - 1次マルコフモデル: 前の状態 $\mathbf{z}_{t-1}$ のみに依存
  - プロセス雑音 $\mathbf{w}_{t-1}$: モデル化誤差などを表現

– 確率密度による表現 $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$

一般的な動的システムのモデル

状態空間モデル: state space model

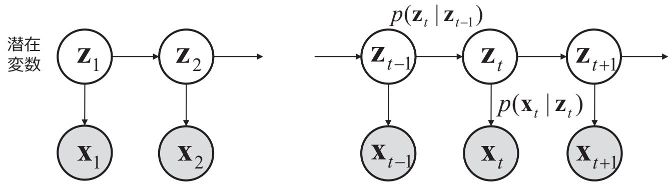
– 観測ベクトル $\mathbf{x}_t$
  - 観測値を要素に持つベクトル
  - 記号(言語)，生のセンサ信号や抽出した特徴量など
    画像+音声などマルチモーダルデータも含む

– 観測方程式 $\mathbf{x}_t = \mathbf{h}_t(\mathbf{z}_t, \mathbf{v}_t)$
  - 状態 $\mathbf{z}_t$ が観測系を通じてどう観測されるかを表現
  - 観測雑音 $\mathbf{v}_t$: 観測系の誤差・雑音を表現

– 確率密度による表現 $p(\mathbf{x}_t \mid \mathbf{z}_t)$

一般的な動的システムのモデル

状態空間モデル: state space model

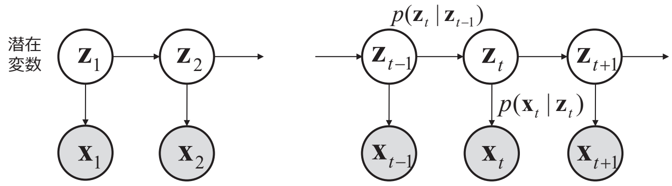


トラッキング・フィルタリングなどの今回の問題設定

**モデル（方程式・分布）は与えられた上で
得られた観測値の時系列 $\mathbf{x}_{1:t} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t]$ から
再帰的に状態 $\mathbf{z}_t$, $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ を推定する問題**

一般的な動的システムのモデル

状態空間モデル: state space model



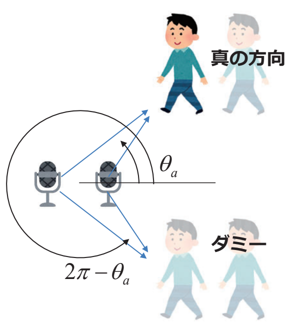
モデル $\mathbf{z}_t = \mathbf{f}_{t-1}(\mathbf{z}_{t-1}, \mathbf{w}_{t-1})$ $\quad p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$ **（のパラメータ）**
$\mathbf{x}_t = \mathbf{h}_t(\mathbf{z}_t, \mathbf{v}_t)$ $\quad p(\mathbf{x}_t \mid \mathbf{z}_t)$

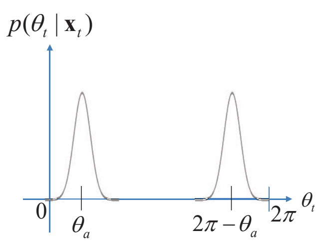
が未知の場合はデータを用いた事前学習も可能

例: $\mathbf{h}, \mathbf{f}$ や密度/分布関数（のパラメータ予測）を Neural Network で実装することも可

基本的に分布を求める

例: 2マイクの場合（到来時間差が対称性）



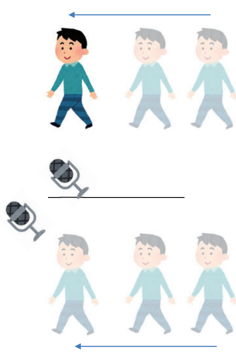


時刻毎に推定するだけでは
曖昧性が残る

基本的に分布を求める

例: 2マイクの場合（到来時間差が対称性）

次の時刻



時間的繋がりを活用する

1. マイク自体の方向を制御
2. 音源の移動予測

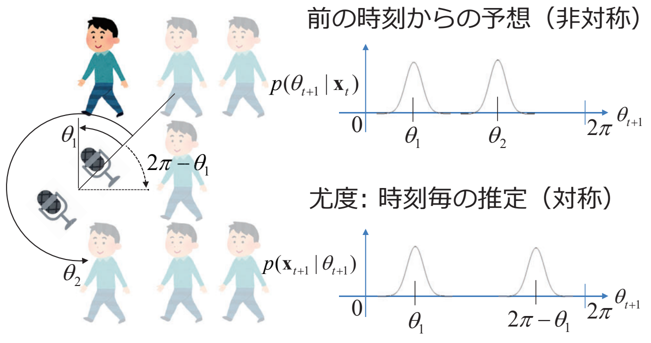
の2点で曖昧性を解消できる

他の情報（vision）で絞り込める場合も
似たような話

（ただし，あくまでも “お話” ）

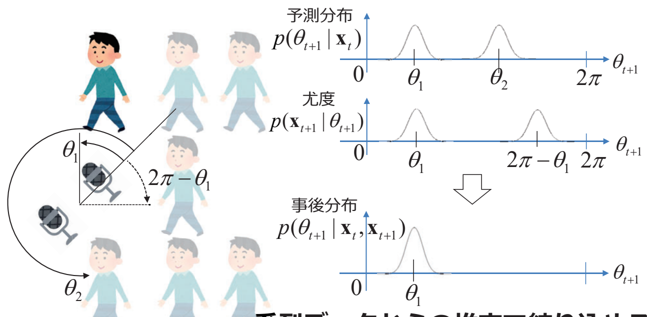
基本的に分布を求める

例: 2マイクの場合（到来時間差が対称性）



基本的に分布を求める

例: 2マイクの場合（到来時間差が対称性）



系列データからの推定で絞り込める
曖昧性(情報)は保持しておく必要性

再帰的な定式化
確率モデル的な視点から

事後分布に基づく定式化

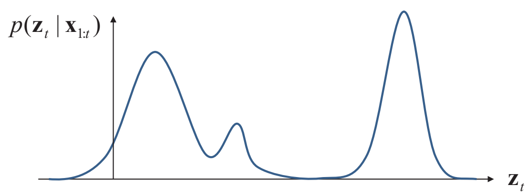
現状態 $\mathbf{z}_t$ の事後分布 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ を求める

– その後に値を１つに決めるなら決める

良さを測る関数に利用可 例: 平均値，最大事後確率，…

$\hat{\mathbf{z}}_t = \int \mathbf{z}_t p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \mathrm{d}\mathbf{z}_t \qquad \hat{\mathbf{z}}_t = \arg\max_{\mathbf{z}_t} p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ …

下記のように山が2つある分布だとどうなる？



事後分布に基づく定式化

現状態 $\mathbf{z}_t$ の事後分布 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ を求める

– その後に値を１つに決めるなら決める

良さを測る関数に利用可 例: 平均値，最大事後確率，…

$\hat{\mathbf{z}}_t = \int \mathbf{z}_t p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \mathrm{d}\mathbf{z}_t \qquad \hat{\mathbf{z}}_t = \arg\max_{\mathbf{z}_t} p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ …

山が2つある分布だとどうなる？

– ここでは**下記の３つだけ**知っていればよい

ベイズの定理 $p(\mathbf{x} \mid \mathbf{y}) = \dfrac{p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \quad \left(= \dfrac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x}} \right)$

加法定理/周辺化 $p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y} = \int p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \mathrm{d}\mathbf{y}$

独立性（依存性） $p(\mathbf{x} \mid \mathbf{y}) = p(\mathbf{x})$ （$\mathbf{x}, \mathbf{y}$ が独立な場合）

事後分布に基づく定式化

現状態 $\mathbf{z}_t$ の事後分布 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ を求める

– その後に値を１つに決めるなら決める

良さを測る関数に利用可 例: 平均値，最大事後確率，…

$\hat{\mathbf{z}}_t = \int \mathbf{z}_t p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \mathrm{d}\mathbf{z}_t \qquad \hat{\mathbf{z}}_t = \arg\max_{\mathbf{z}_t} p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ …

山が2つある分布だと？

– 以降，よくわからなくなったら，**積分を和に変換して離散的な場合をイメージ**してみること

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y} = \int p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \mathrm{d}\mathbf{y}$$

$$\Rightarrow p(\mathbf{x}) = \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{y}} p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y})$$

あらゆる y (例えばサイコロの目)を列挙して，確率を足し合わせる

事後分布に基づく定式化

事後分布(データを見た後)の再帰的表現を導く

$$p(\underline{\mathbf{z}_t}\,|\,\underline{\mathbf{x}_{1:t}}) = p(\mathbf{z}_t\,|\,\mathbf{x}_t,\mathbf{x}_{1:t-1})$$

ベイズの定理
（$\mathbf{x}_{1:t-1}$ は共通）

動かして調べたいもの

観測してすでに知っている値（固定値）

$$= \frac{p(\mathbf{x}_t\,|\,\mathbf{z}_t,\mathbf{x}_{1:t-1})\,p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1})}{p(\mathbf{x}_t\,|\,\mathbf{x}_{1:t-1})}$$

予測と観測のAND

予測がどれだけ当たっているか

– モデルの仮定より

尤度 $p(\mathbf{x}_t\,|\,\mathbf{z}_t,\mathbf{x}_{1:t-1}) = p(\mathbf{x}_t\,|\,\mathbf{z}_t)$

x: 固定値 → 尤度関数
データがどの程度 z の設定に当てはまってるかを表現

– 事前/予測分布: t-1 までのデータに基づく予測
（時刻 t のデータを見る前）

$$p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1})$$

事後分布に基づく定式化

– 事前/予測分布（データを見る前）の展開

$$p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1}) = \int p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1},\mathbf{x}_{1:t-1})\,p(\mathbf{z}_{t-1}\,|\,\mathbf{x}_{1:t-1})\mathrm{d}\mathbf{z}_{t-1}$$

ある同時分布を周辺化した分布だと仮定

$$= \int p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})\,\underline{p(\mathbf{z}_{t-1}\,|\,\mathbf{x}_{1:t-1})}\mathrm{d}\mathbf{z}_{t-1}$$

– モデルの仮定（1次マルコフ）より

状態遷移確率 $p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1},\mathbf{x}_{1:t-1}) = p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})$

– **t-1 時刻の事後分布: 既に求まっている** （初期値は必要）

$$p(\mathbf{z}_{t-1}\,|\,\mathbf{x}_{1:t-1})$$

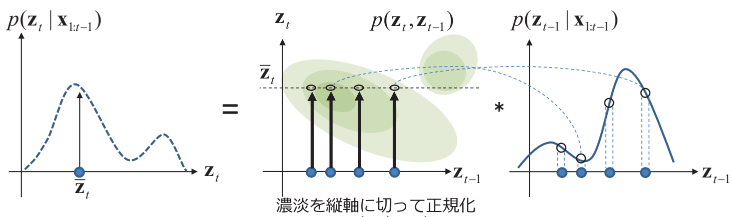
事後分布に基づく定式化

– 補足: 事前/予測分布の展開部分

$$p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1}) = \int \overset{\text{状態遷移}}{p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})}\,p(\mathbf{z}_{t-1}\,|\,\mathbf{x}_{1:t-1})\mathrm{d}\mathbf{z}_{t-1}$$

$\mathbf{z}_t$ を生じるパターンを列挙して sum (全パタン場合分け)

1つの点 $\bar{\mathbf{z}}_t$ を取って，1つ1つ平易に考えると…



濃淡を縦軸に切って正規化
$= p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})$

事後分布に基づく定式化

まとめ: 事後分布の再帰的な表現

動かして調べたいもの　すでに知っている値（固定値）　積（データとのマッチ具合でとったAND的条件）

$$p(\underline{\mathbf{z}_t}\,|\,\underline{\mathbf{x}_{1:t}}) = \frac{p(\mathbf{x}_t\,|\,\mathbf{z}_t)\,p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1})}{p(\mathbf{x}_t\,|\,\mathbf{x}_{1:t-1})} \quad \left(= \int p(\mathbf{x}_t|\mathbf{z}_t)p(\mathbf{z}_t|\mathbf{x}_{1:t-1})\mathrm{d}\mathbf{z}_t\right)$$

$$p(\mathbf{z}_t\,|\,\mathbf{x}_{1:t-1}) = \int p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})\,p(\mathbf{z}_{t-1}\,|\,\mathbf{x}_{1:t-1})\mathrm{d}\mathbf{z}_{t-1}$$

$\mathbf{z}_t$ を生じるパターンを列挙して sum (全パタン場合分け)

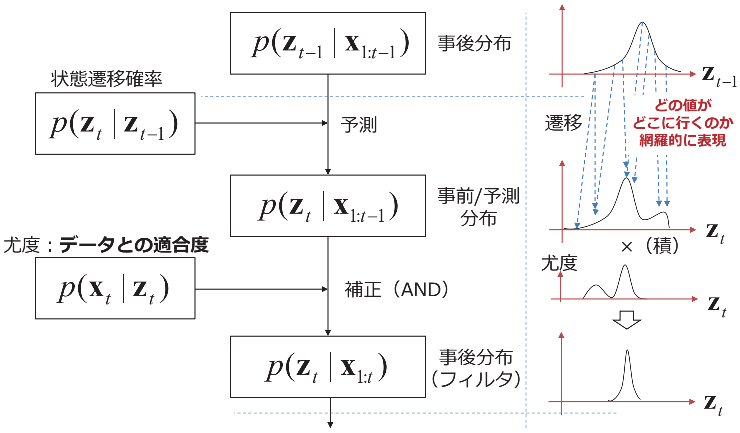
(離散確率変数の場合: 積分 → 和)

状態遷移確率 $p(\mathbf{z}_t\,|\,\mathbf{z}_{t-1})$

尤度 $p(\mathbf{x}_t\,|\,\mathbf{z}_t)$

以降（講義の後半）は，**モデルが既知**だと仮定して話を進める

予測と事後分布の更新の流れ



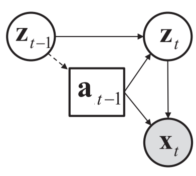
部分観測マルコフ決定過程: POMDP
Partially observed Markov decision process

システムの"行動" が加わったモデル

– POMDPとして第7回スライドにちらっと登場

– 自身の行動 a → 自分で決める = ランダムではない（自分のポリシーに依存）

– 行動を決める関数 g: 別のコスト関数で学習



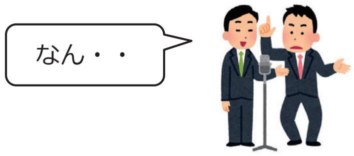
$$a_{t-1} \leftarrow g(\mathbf{z}_{t-1})$$

いわゆる"信念分布"

$$p(\mathbf{z}_t|\mathbf{x}_{1:t},a_{1:t-1}) = \frac{p(\mathbf{x}_t|\mathbf{z}_t,a_{t-1})\,p(\mathbf{z}_t|\mathbf{x}_{1:t-1},a_{1:t-1})}{p(\mathbf{x}_t|a_{1:t-1},\mathbf{x}_{1:t-1})}$$

$$p(\mathbf{z}_t|\mathbf{x}_{1:t-1},a_{1:t-1}) = \int p(\mathbf{z}_t|\mathbf{z}_{t-1},a_{t-1})\,p(\mathbf{z}_{t-1}|\mathbf{x}_{1:t-1},a_{1:t-2})\mathrm{d}\mathbf{z}_{t-1}$$

考える例えは何でもよい: 発音の推測



予測: 次なんて言いそうか
- 「なん」→「で」→「やねん」
- 「なん」→「じ」→「かな」

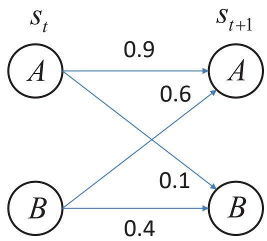
データ（聴覚情報）：「で」っぽく聞こえた
事後・予測: 「やねん」が続くことを確信

動きの予測（"相手はこう考えて動くだろう"），フェイントの掛け合い，
など実世界でも似たようなことしているはず
「場合分けによる予想＋実際に起こったことによる絞り込み」という過程

Mini Quiz #1

$P(s_t = A) = 0.3$

$P(s_t = B) = 0.7$



$P(s_{t+1} = A \mid s_t = A) = 0.9$

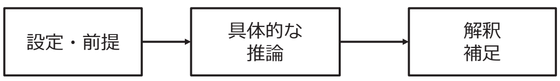
$P(s_{t+1} = B \mid s_t = A) = 0.1$

$P(s_{t+1} = A \mid s_t = B) = 0.6$

$P(s_{t+1} = B \mid s_t = B) = 0.4$

- 上記の設定の時，時刻 t+1 における各状態の確率 $P(s_{t+1} = A)$ $P(s_{t+1} = B)$ を下記の式に従って計算せよ

$$P(s_{t+1}) = \sum_{s_t} P(s_{t+1} \mid s_t) P(s_t)$$



KALMAN FILTER
確率モデル的な視点から

他: 現代制御理論的な視点

カルマンフィルタ

線形・ガウスの制約下でのモデル, $\mathbf{z}_t$ の推定
- プロセス・観測方程式は線形
- 雑音 $\mathbf{v}_t, \mathbf{w}_t$ はガウス分布に従う

プロセス $\mathbf{z}_t = \mathbf{A}_{t-1}\mathbf{z}_{t-1} + \mathbf{w}_{t-1}$ $\quad \mathbf{w} \sim N(\mathbf{w} \mid \mathbf{0}, \boldsymbol{\Gamma})$

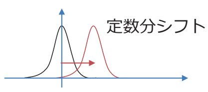
観測 $\mathbf{x}_t = \mathbf{C}_t\mathbf{z}_t + \mathbf{v}_t$ $\quad \mathbf{v} \sim N(\mathbf{v} \mid \mathbf{0}, \boldsymbol{\Sigma})$

初期値 $\mathbf{z}_1 = \boldsymbol{\mu}_0 + \mathbf{u}$ $\quad \mathbf{u} \sim N(\mathbf{u} \mid \mathbf{0}, \mathbf{P}_1)$

$\mathbf{A}_t, \mathbf{C}_t, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\mu}_0, \mathbf{P}_1$ は既知のパラメータ
$\boldsymbol{\Sigma}, \boldsymbol{\Gamma}$ に t がついてても可

カルマンフィルタ

線形・ガウスの制約下でのモデル, $\mathbf{z}_t$ の推定
- プロセス・観測方程式は線形
- 雑音 $\mathbf{v}_t, \mathbf{w}_t$ はガウス分布に従う



プロセス $p(\mathbf{z}_t \mid \mathbf{z}_{t-1}) = N(\mathbf{z}_t \mid \mathbf{A}_{t-1}\mathbf{z}_{t-1}, \boldsymbol{\Gamma})$

$\underbrace{\mathbf{z}_t = \mathbf{A}_{t-1}\mathbf{z}_{t-1}}_{\text{given(定数)}} + \underbrace{\mathbf{w}_{t-1}}_{\text{ランダム}}$

観測 $p(\mathbf{x}_t \mid \mathbf{z}_t) = N(\mathbf{x}_t \mid \mathbf{C}_t\mathbf{z}_t, \boldsymbol{\Sigma})$

$\underbrace{\mathbf{x}_t = \mathbf{C}_t\mathbf{z}_t}_{\text{given(定数)}} + \underbrace{\mathbf{v}_t}_{\text{ランダム}}$

初期値 $p(\mathbf{z}_1) = N(\mathbf{z}_1 \mid \boldsymbol{\mu}_0, \mathbf{P}_1)$

$\mathbf{A}_t, \mathbf{C}_t, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\mu}_0, \mathbf{P}_1$ は既知のパラメータ

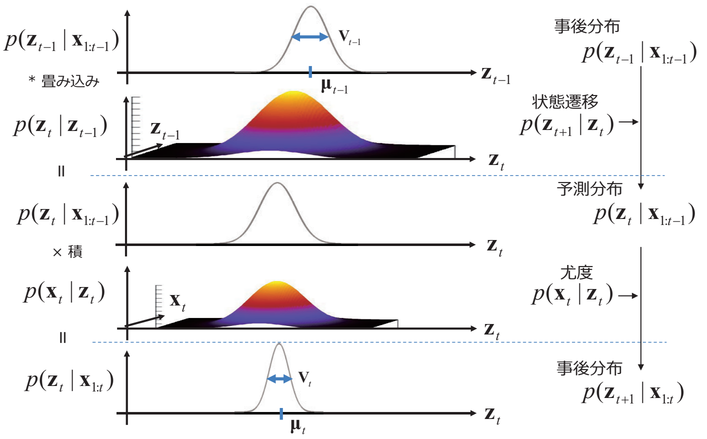
再帰的な推定

線形・ガウスモデル
- すべての潜在変数と観測変数にわたる同時分布は**ガウス分布**
  - 尤度・予測/事前分布・事後分布がガウス分布
    ＝分布のピークは１つだけ

平均（一番それっぽい値）と**分散**（バラつき）の2つを**再帰的に更新できれば良い**
（ガウス分布は平均・分散で形状が決まるから）

- 多変量ガウス分布の周辺分布・条件付き分布の関係を利用して導出

カルマンフィルタの流れ



ガウス分布の周辺分布と条件付き分布

次の **2つ** を知っていれば式は追える

下記が与えられたとき

$$p(\mathbf{x}) = N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$

$$p(\mathbf{y}|\mathbf{x}) = N(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{L}^{-1})$$

$\mathbf{y}$ の周辺分布と $\mathbf{x}$ の条件付き分布は

$$p(\mathbf{y}) = N(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{L}^{-1}+\mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}) \quad \text{(A)}$$

$$p(\mathbf{x}|\mathbf{y}) = N(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y}-\mathbf{b})+\boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \quad \text{(B)}$$

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}$$

推論

流れ: 事後分布から平均・分散の関係を追う

– 時刻 t の事後分布（ガウス）のパラメータ設定

$$p(\mathbf{z}_t|\mathbf{x}_{1:t}) = N(\mathbf{z}_t|\underline{\boldsymbol{\mu}_t, \mathbf{V}_t})$$

– 下記の再帰式に基づき時刻 t-1との関係を導出

$$p(\mathbf{z}_t|\mathbf{x}_{1:t}) = \frac{p(\mathbf{x}_t|\mathbf{z}_t)p(\mathbf{z}_t|\mathbf{x}_{1:t-1})}{p(\mathbf{x}_t|\mathbf{x}_{1:t-1})}$$

$$p(\mathbf{z}_t|\mathbf{x}_{1:t-1}) = \int p(\mathbf{z}_t|\mathbf{z}_{t-1})p(\mathbf{z}_{t-1}|\mathbf{x}_{1:t-1})d\mathbf{z}_{t-1}$$

推論

– とりあえずひらたく展開

$$\underbrace{p(\mathbf{z}_t|\mathbf{x}_{1:t})}_{\text{事後分布}} = \frac{\overbrace{p(\mathbf{x}_t|\mathbf{z}_t)}^{\text{尤度}}\overbrace{p(\mathbf{z}_t|\mathbf{x}_{1:t-1})}^{\text{予測分布}}}{p(\mathbf{x}_t|\mathbf{x}_{1:t-1})}$$

$$= \frac{1}{p(\mathbf{x}_t|\mathbf{x}_{1:t-1})}p(\mathbf{x}_t|\mathbf{z}_t)\int \overbrace{p(\mathbf{z}_t|\mathbf{z}_{t-1})}^{\text{状態遷移確率}}\underbrace{p(\mathbf{z}_{t-1}|\mathbf{x}_{1:t-1})}_{\text{t-1時刻の事後分布}}d\mathbf{z}_{t-1}$$

$$p(\mathbf{z}_{t-1}|\mathbf{x}_{1:t-1}) = N(\mathbf{z}_{t-1}|\underline{\boldsymbol{\mu}_{t-1}, \mathbf{V}_{t-1}})$$

事後分布パラメータ $\boldsymbol{\mu}_{t-1}, \mathbf{V}_{t-1}$ と $\boldsymbol{\mu}_t, \mathbf{V}_t$ をつなげる

$\boldsymbol{\mu}_{t-1}, \mathbf{V}_{t-1}$ から出発して順番に計算していく

<u>計算過程もそこそこ重要だが，結果の解釈が大切</u>

推論

予測分布（データを見る前）の平均・共分散行列

$$p(\mathbf{z}_t|\mathbf{x}_{1:t-1}) = \int \overbrace{p(\mathbf{z}_t|\mathbf{z}_{t-1})}^{\text{状態遷移確率}}\overbrace{p(\mathbf{z}_{t-1}|\mathbf{x}_{1:t-1})}^{\text{t-1時刻の事後分布}}d\mathbf{z}_{t-1}$$

分布のピーク値

$$= \int N(\mathbf{z}_t|\mathbf{A}_{t-1}\mathbf{z}_{t-1}, \boldsymbol{\Gamma})N(\mathbf{z}_{t-1}|\boxed{\boldsymbol{\mu}_{t-1}}, \mathbf{V}_{t-1})d\mathbf{z}_{t-1}$$

式（A）を適用

$$= N(\mathbf{z}_t|\mathbf{A}_{t-1}\boldsymbol{\mu}_{t-1}, \mathbf{P}_t)$$

ただし

プロセス方程式に基づく状態の予測

$$\hat{\mathbf{z}}_t = \mathbf{A}_{t-1}\boldsymbol{\mu}_{t-1}$$ ピークの点（平均）：そのまま変換

$$\mathbf{P}_t = \mathbf{A}_{t-1}\mathbf{V}_{t-1}\mathbf{A}_{t-1}^{\mathrm{T}} + \boldsymbol{\Gamma}$$ 雑音の分散が加算（再生性からも）

$$\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t^T = \mathbf{A}_{t-1}(\boldsymbol{\mu}_{t-1}\boldsymbol{\mu}_{t-1}^{\mathrm{T}})\mathbf{A}_{t-1}^{\mathrm{T}}$$

推論

事後分布（データを見た後）の平均・共分散行列

$$p(\mathbf{z}_t|\mathbf{x}_{1:t}) = \frac{p(\mathbf{x}_t|\mathbf{z}_t)p(\mathbf{z}_t|\mathbf{x}_{1:t-1})}{p(\mathbf{x}_t|\mathbf{x}_{1:t-1})}$$

予測分布 $p(\mathbf{z}_t|\mathbf{x}_{1:t-1}) = N(\mathbf{z}_t|\underline{\mathbf{A}_{t-1}\boldsymbol{\mu}_{t-1}}, \mathbf{P}_t)$

尤度 $p(\mathbf{x}_t|\mathbf{z}_t) = N(\mathbf{x}_t|\mathbf{C}_t\mathbf{z}_t, \boldsymbol{\Sigma})$

式（B）を適用

$$\boldsymbol{\mu}_t = \mathbf{V}_t(\mathbf{C}_t^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_t + \mathbf{P}_t^{-1}\underline{\mathbf{A}_{t-1}\boldsymbol{\mu}_{t-1}})$$

$$\mathbf{V}_t = (\mathbf{P}_t^{-1} + \mathbf{C}_t^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{C}_t)^{-1}$$

$\hat{\mathbf{z}}_t$ 先ほどの予測値（ピークの点）

一応繋がったが，もう一つよくわからないので整理を続ける

推論

– 共分散行列の整理（逆行列部分が対象）

$$\mathbf{V}_t = (\mathbf{P}_t^{-1} + \mathbf{C}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{C}_t)^{-1}$$

スライド下の I) より

$$= \mathbf{P}_t - \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} (\boldsymbol{\Sigma} + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}})^{-1} \mathbf{C}_t \mathbf{P}_t$$

カルマンゲイン $\mathbf{K}_t$ を（天下り的に）定義

$$\mathbf{K}_t = \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} (\boldsymbol{\Sigma} + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}})^{-1}$$

$$\mathbf{V}_t = \mathbf{P}_t - \mathbf{K}_t \mathbf{C}_t \mathbf{P}_t$$

$$= (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \mathbf{P}_t$$

I) 逆行列の補助定理 $(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$

推論

– 平均を（ゴリゴリ）整理

$$\boldsymbol{\mu}_t = \mathbf{V}_t (\mathbf{C}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{x}_t + \mathbf{P}_t^{-1} \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1})$$

Vをかっこの中にかけて
前ページの $\mathbf{V}_t = (\mathbf{P}_t^{-1} + \mathbf{C}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{C}_t)^{-1}$

$$= [(\mathbf{P}_t^{-1} + \mathbf{C}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{C}_t)^{-1} \mathbf{C}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}] \mathbf{x}_t + \mathbf{V}_t \mathbf{P}_t^{-1} \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}$$

スライド下の II) より

$$= \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} (\mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} + \boldsymbol{\Sigma})^{-1} \mathbf{x}_t + \mathbf{V}_t \mathbf{P}_t^{-1} \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}$$

第1項: カルマンゲインの定義
第2項: 全ページ下の V を適用

$$= \mathbf{K}_t \mathbf{x}_t + (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}$$

$$= \underbrace{\mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}}_{\text{予測値 } \hat{\mathbf{z}}_t} + \mathbf{K}_t \overbrace{(\mathbf{x}_t - \underbrace{\mathbf{C}_t}_{\text{観測方程式}} \underbrace{\mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}}_{\text{予測値 } \hat{\mathbf{z}}_t})}^{\text{観測値との誤差（イノベーション）}}$$

観測値のシミュレート
+
誤差に基づき
状態の予測値を補正

II) 逆行列に関する等式 $(\mathbf{P}^{-1} + \mathbf{B}^T \mathbf{R}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^T (\mathbf{B}\mathbf{P}\mathbf{B}^T + \mathbf{R})^{-1}$

推論

まとめ: 下記の再帰計算で事後分布がわかる

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = N(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \mathbf{V}_t)$$

$$\mathbf{P}_t = \mathbf{A}_{t-1} \mathbf{V}_{t-1} \mathbf{A}_{t-1}^{\mathrm{T}} + \boldsymbol{\Gamma}$$

$$\mathbf{K}_t = \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} (\boldsymbol{\Sigma} + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}})^{-1}$$

$$\boldsymbol{\mu}_t = \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1} + \mathbf{K}_t (\mathbf{x}_t - \mathbf{C}_t \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1})$$

$$\mathbf{V}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \mathbf{P}_t$$

初期値

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{K}_1 (\mathbf{x}_1 - \mathbf{C}_1 \boldsymbol{\mu}_0)$$
$$\mathbf{V}_1 = (\mathbf{I} - \mathbf{K}_1 \mathbf{C}_1) \mathbf{P}_1$$
$$\mathbf{K}_1 = \mathbf{P}_1 \mathbf{C}_1^{\mathrm{T}} (\boldsymbol{\Sigma} + \mathbf{C}_1 \mathbf{P}_1 \mathbf{C}_1^{\mathrm{T}})^{-1}$$

$\mathbf{P}_1$ パラメータ

処理の見通しをよくするためもう少し整理

推論

分母の正規化定数 $p(\mathbf{x}_t \mid \mathbf{x}_{1:t-1})$ も見ておく

– 同様にガウス分布

$$p(\mathbf{x}_t \mid \mathbf{x}_{1:t-1}) = N(\mathbf{x}_t \mid \mathbf{C}_t \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma} + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}})$$

– **モデルがデータをどれだけ予測できたかを表現**

- **値が低い＝予測が外れている（低信頼度）**

– **平均・分散: 観測方程式に基づいて予測したもの**

$$\hat{\mathbf{x}}_t = \underbrace{\mathbf{C}_t}_{\text{観測行列}} \underbrace{\mathbf{A}_{t-1}}_{\text{状態遷移}} \underbrace{\boldsymbol{\mu}_{t-1}}_{\text{1時刻前のピーク値}} \qquad \mathbf{S}_t = \underbrace{\boldsymbol{\Sigma}}_{\text{観測雑音の分散}} + \underbrace{\mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}}}_{\text{予測値の分散}}$$

$\mathbf{P}_t = \mathbf{A}_{t-1} \mathbf{V}_{t-1} \mathbf{A}_{t-1}^{\mathrm{T}} + \boldsymbol{\Gamma}$

これらの値: 平均更新とゲイン計算で間接的に利用

$$\boldsymbol{\mu}_t = \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1} + \mathbf{K}_t (\mathbf{x}_t - \hat{\mathbf{x}}_t) \qquad \mathbf{K}_t = \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}} \mathbf{S}_t^{-1}$$

状態と観測の推定分散の比
（誤差の信頼度）のようなもの

ステップの分解と意味の解釈

状態予測 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$

– 平均: 移動 $\quad \hat{\mathbf{z}}_t = \mathbf{A}_{t-1} \boldsymbol{\mu}_{t-1}$

– 分散 $\quad \mathbf{P}_t = \underbrace{\mathbf{A}_{t-1} \mathbf{V}_{t-1} \mathbf{A}_{t-1}^{\mathrm{T}}}_{\text{過程を経た分}} + \underbrace{\boldsymbol{\Gamma}}_{\text{誤差・雑音分は加算 不確定度合いが上がる}}$

モデルに従って
状態の
平均・分散を予測

データ予測 $p(\mathbf{x}_t \mid \mathbf{x}_{1:t-1})$

– 平均 $\quad \hat{\mathbf{x}}_t = \mathbf{C}_t \hat{\mathbf{z}}_t$

– 分散 $\quad \mathbf{S}_t = \underbrace{\mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^{\mathrm{T}}}_{\text{過程を経た分}} + \underline{\boldsymbol{\Sigma}}$

モデルに従って
観測データの
平均・分散を予測

フィルタ $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$

– ゲイン $\quad \mathbf{K}_t = \mathbf{P}_t \mathbf{C}_t^T \mathbf{S}_t^{-1}$

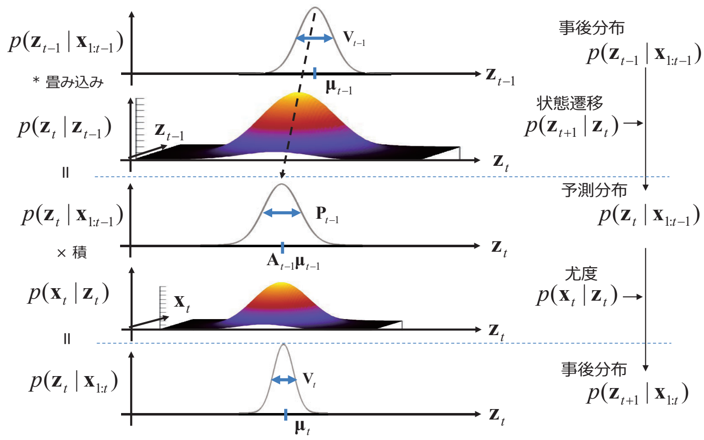
予測の曖昧性で補正
状態の分散が小さいほど
観測の分散が大きいほど　小さい

– 平均: 補正 $\quad \boldsymbol{\mu}_t = \hat{\mathbf{z}}_t + \mathbf{K}_t \underline{(\mathbf{x}_t - \hat{\mathbf{x}}_t)}$

– 分散: 補正 $\quad \mathbf{V}_t = \mathbf{P}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^{\mathrm{T}}$

予測誤差に
基づいて補正

カルマンフィルタの流れ



Kalman Filter の学習

観測データからモデルパラメータを学習

フィルタ: データ + パラメータありで状態を推定

学習: データのみでパラメータ・状態を推定

– 伝達関数と音源信号の両方が未知など

cf. エコーキャンセラ

教師なし学習（ブラインド推定）の問題

– 勾配法，EMアルゴリズムなどを用いて推定

- 局所解の存在に注意 = 初期値依存性

他: Neural network - variational auto encoder

他: 状態にパラメータ自体を含める考え方

非線形関数への対応

Extended Kalman Filter

– 各方程式を1次のテーラー級数展開で局所的に線形化

- 平均値は非線形関数を利用 $\hat{\boldsymbol{\mu}}_t = \mathbf{f}_{t-1}(\boldsymbol{\mu}_{t-1})$ $\hat{\mathbf{x}}_t = \mathbf{h}_t(\hat{\boldsymbol{\mu}}_t)$
- 共分散計算でヤコビ行列を利用 $\hat{\mathbf{A}}_{t-1}, \hat{\mathbf{C}}_t$

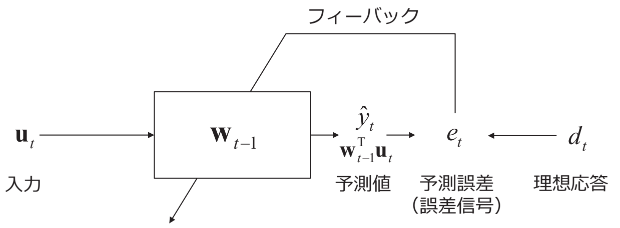
Unscented Kalman Filter

– EKF は平均値が非線形性に反映

– シグマポイントと呼ばれるサンプル点を用いて，平均値と共分散行列を再構成

実際に応用するなら，UKF まで試して性能比較した方が良い

再掲: LMS の挙動



LMS $\quad \mathbf{w}_t = \mathbf{w}_{t-1} + 2\mu(d_t - \hat{y}_t)\mathbf{u}_t$

Mini Quiz #2 (余裕があれば)

- エコーキャンセラのフィルタをカルマンフィルターの枠組みで推定できる．プロセス方程式，観測方程式をどのように当てはめればよいだろうか

  – 応用する際，柔軟に型にはめられるかどうか

NLSMの例

$$\min_{\mathbf{w}_t} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|^2 \quad \overset{?}{\Rightarrow} \quad \mathbf{z}_t = \mathbf{A}_{t-1}\mathbf{z}_{t-1} + \mathbf{w}_t$$

$$y_t = \mathbf{w}_t^T \mathbf{u}_t \quad \overset{?}{\Rightarrow} \quad \mathbf{x}_t = \mathbf{C}_t\mathbf{z}_t + \mathbf{v}_t$$

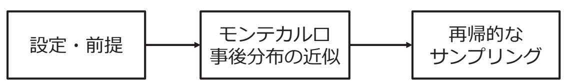
Mini Quiz #3 (余裕があれば)

- カルマンフィルタ（だけではないが）では，初期値 $\boldsymbol{\mu}_0, \mathbf{P}_1$ や雑音/誤差項の共分散パラメータ $\boldsymbol{\Sigma}, \boldsymbol{\Gamma}$ の設定も重要となる.

- 下記を考えてみましょう

  – 初期値が検討外れの場合、どういう挙動をしそうか

  – 対象を正しくモデル化できていない場合、共分散パラメータをどのように設定するのがよさそうか



PARTICLE FILTER

パーティクルフィルタ（粒子フィルタ）

非線形・非ガウスへの対応

→ 有限サンプルで分布を近似
モンテカルロ法の導入

☹ Kalman Filter: ガウス分布（ピークが1つ）

– 各サンプル（粒子/仮説）は過程をシミュレート
  - 確率的に決まる箇所も乱数生成して "値" を計算

SIR (Sampling Important Resampling)
アルゴリズムを天下り的にざっと見る
・広く使われている実装
・外れ値に弱い ⇔ 実装が容易

パーティクルフィルタ（粒子フィルタ）

下記は既知だと仮定

– プロセス方程式 もしくは 状態遷移確率

$$\mathbf{z}_t = \mathbf{f}_{t-1}(\mathbf{z}_{t-1}, \mathbf{w}_{t-1}) \quad \mathbf{w}_{t-1} \sim p(\mathbf{w}_{t-1})$$

$$p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$$

– 観測方程式 もしくは 尤度

$$\mathbf{x}_t = \mathbf{h}_t(\mathbf{z}_t, \mathbf{v}_t) \qquad \mathbf{v}_t \sim p(\mathbf{v}_t)$$

$$p(\mathbf{x}_t \mid \mathbf{z}_t)$$

補足: "過程をシミュレート" 状態遷移における例

手元にある情報（前提）

– モデルとそのパラメータ，および，$\mathbf{z}_{t-1}$

確率モデルの場合

1. $\mathbf{z}_t$ を1つサンプリング: $\mathbf{z}_t \sim p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$

おわり

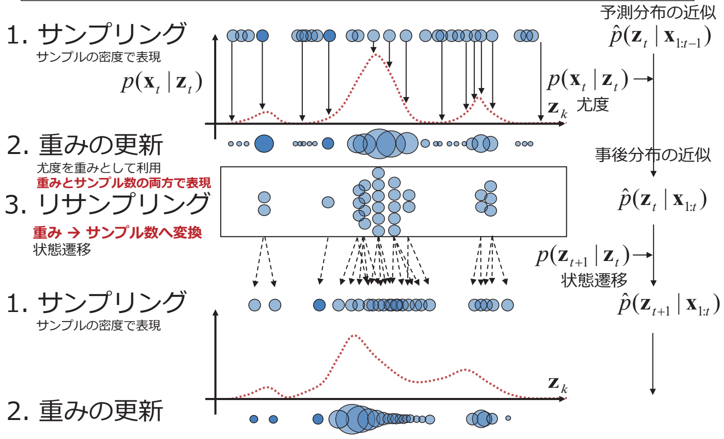
方程式の場合 $\mathbf{z}_t = \mathbf{f}_{t-1}(\mathbf{z}_{t-1}) + \mathbf{w} \qquad \mathbf{w} \sim p(\mathbf{w})$

1. $\mathbf{f}_{t-1}(\mathbf{z}_{t-1})$ を計算: $\mathbf{f}$ は確定的に決まる関数
2. $\mathbf{w}$ を1つサンプリング: $\mathbf{w} \sim p(\mathbf{w})$
3. 1と2の結果を足す

おわり

詳細はモデル依存

SIR フィルタの流れ



モンテカルロ法

積分の計算

– 解析的な計算が難しい

$$E[f(\mathbf{x})] = \int f(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

– $p(\mathbf{x})$ に従う L 個のサンプル $\{\mathbf{x}^l; l = 1, \ldots, L\}$ で近似

$$\hat{p}(\mathbf{x}) = \frac{1}{L} \sum_l \delta(\mathbf{x} - \mathbf{x}^{(l)})$$

$$E[f(\mathbf{x})] \approx \frac{1}{L} \sum_l f(\mathbf{x}^{(l)})$$ サンプルそれぞれが $\frac{1}{L}$ の重みをもつ

※ 関数 f が確率分布（尤度関数）の場合もある $p(\mathbf{x}) = \int p(\mathbf{x} \mid \mathbf{y}) p(\mathbf{y}) \mathrm{d}\mathbf{y}$

事後分布による期待値の近似

観測値 $\mathbf{x}_{1:t}$ が与えられた元で、事後分布 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ からの L 個のサンプルを得たい

ひとまず任意の関数 f の事後分布による期待値がどうなるかを見る

f 以外の部分が事後分布部分

$$E[f(\mathbf{z}_t)] = \int f(\mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \mathrm{d}\mathbf{z}_t$$

ベイズの定理

$$= \int f(\mathbf{z}_t) \frac{p(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})}{p(\mathbf{x}_t \mid \mathbf{x}_{1:t-1})} \mathrm{d}\mathbf{z}_t$$

モデルの仮定

$$= \frac{\int f(\mathbf{z}_t) p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}{\int p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}$$

事後分布による期待値の近似

観測値 $\mathbf{x}_{1:t}$ が与えられた元で、事後分布 $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ からの L 個のサンプルを得たい

$$E[f(\mathbf{z}_t)] = \frac{\int f(\mathbf{z}_t) p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}{\int p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t} \longleftarrow \text{正規化定数}$$

$$\approx \frac{\frac{1}{L}\sum_l f(\mathbf{z}_t^{(l)}) p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\frac{1}{L}\sum_{l'} p(\mathbf{x}_t \mid \mathbf{z}_t^{(l')})} \longleftarrow \text{定数}$$

予測分布からのサンプル
$\mathbf{z}_t^{(l)} \sim p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$

尤度による重み付き平均
※ サンプルは
予測分布に従ったもの

$$= \sum_{l=1}^{L} w_t^{(l)} f(\mathbf{z}_t^{(l)}) \qquad w_t^{(l)} = \frac{p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\sum_{m=1}^{L} p(\mathbf{x}_t \mid \mathbf{z}_t^{(m)})}$$

事後分布による期待値の近似

– 期待値計算

$$\mathbf{z}_t^{(l)} \sim p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$$

$$E[f(\mathbf{z}_t)] \approx \sum_{l=1}^{L} w_t^{(l)} f(\mathbf{z}_t^{(l)})$$

サンプル点に
尤度が重みとして加わる

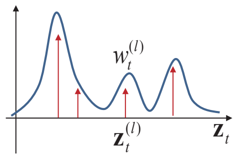
$$w_t^{(l)} = \frac{p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\sum_{m=1}^{L} p(\mathbf{x}_t \mid \mathbf{z}_t^{(m)})}$$

上式: 事後分布を 予測分布からの **L 個のサンプル集合** $\{\mathbf{z}_t^{(l)}\}$ と 対応する**重み** $\{w_t^{(l)}\}$ で離散的に表現

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \sum_{l=1}^{L} w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)})$$

※ 等しい重み $\frac{1}{L}$ ではない

$$\left(\int f(\mathbf{z}_t) \overbrace{\sum_{l=1}^{L} w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)})}^{p(\mathbf{z}_t \mid \mathbf{x}_{1:t})} \mathrm{d}\mathbf{z}_t = \sum_{l=1}^{L} w_t^{(l)} f(\mathbf{z}_t^{(l)}) \right)$$



予測分布の再帰的な表現

逐次的なサンプリング

– 時刻 t で重みとサンプル集合はある $\{\mathbf{z}_t^{(l)}, w_t^{(l)}\}$

– 時刻 t + 1 での重みとサンプルを求める

$$p(\mathbf{z}_{t+1} \mid \mathbf{x}_{1:t}) = \int p(\mathbf{z}_{t+1} \mid \mathbf{z}_t, \mathbf{x}_{1:t}) p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \mathrm{d}\mathbf{z}_t$$

ベイズの定理

$$= \int p(\mathbf{z}_{t+1} \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_t, \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t$$

モデルの仮定

$$= \frac{\int p(\mathbf{z}_{t+1} \mid \mathbf{z}_t) p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}{\int p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}$$

予測分布の再帰的な表現

逐次的なサンプリング

– 時刻 t で重みとサンプル集合はある $\{\mathbf{z}_t^{(l)}, w_t^{(l)}\}$

– 時刻 t + 1 での重みとサンプルを求める

3ページ前で言う f

$$p(\mathbf{z}_{t+1} \mid \mathbf{x}_{1:t}) = \frac{\int p(\mathbf{z}_{t+1} \mid \mathbf{z}_t) p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}{\int p(\mathbf{x}_t \mid \mathbf{z}_t) p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \mathrm{d}\mathbf{z}_t}$$

重み×状態遷移確率

$$\approx \sum_l w_t^{(l)} p(\mathbf{z}_{t+1} \mid \mathbf{z}_t^{(l)}) \qquad \mathbf{z}_t^{(l)} \sim p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$$

$$w_t^{(l)} = \frac{p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\sum_{m=1}^{L} p(\mathbf{x}_t \mid \mathbf{z}_t^{(m)})}$$

予測分布の再帰的な表現

1. 現時刻のサンプルで重み（重要度）計算

$$\mathbf{z}_t^{(l)} \sim p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$$ （すでにサンプルはあると仮定）

$$w_t^{(l)} = \frac{p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\sum_{m=1}^{L} p(\mathbf{x}_t \mid \mathbf{z}_t^{(m)})} \qquad \left(p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \sum_l w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)}) \right)$$

2. 近似した予測分布からのサンプリング
   = **混合分布**からのサンプリング

$$p(\mathbf{z}_{t+1} \mid \mathbf{x}_{1:t}) \approx \sum_l w_t^{(l)} p(\mathbf{z}_{t+1} \mid \mathbf{z}_t^{(l)}) \qquad w_t^{(l)}\text{: 混合重み}$$

2 ができれば，再帰的に 1.重み計算 へ移れる

混合分布からのサンプリング（ステップを分割）

1. どの分布から生成されるか（クラス）を表す変数 l_k をサンプリング（リサンプリング）

$$l_k \sim \mathrm{Cat}(l \mid w_t^{(1)}, \ldots, w_t^{(L)}) \qquad \sum_l w_t^{(l)} = 1 \quad \text{カテゴリカル分布}$$

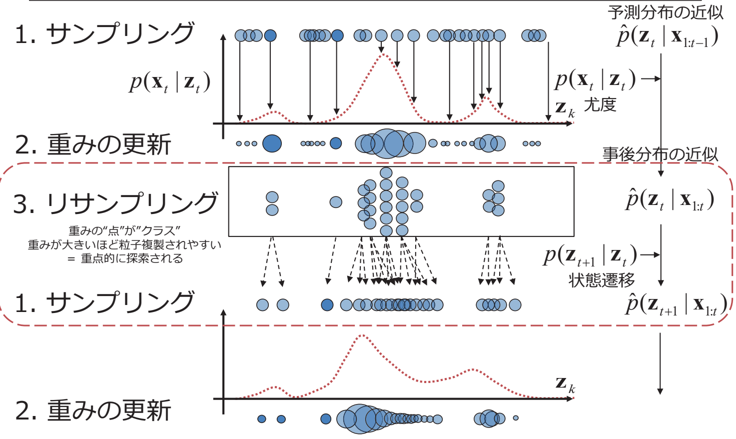
もしくは $\mathbf{z}_t^{(l_k)} \sim \sum_{l=1}^{L} w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)}) \approx p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$

- 事後分布の観点: サンプルの複製手続き
- “重み” → “数”へ変換 & 不要粒子除去: 重みは $1/L$ に

2. 該当する状態遷移分布から $\mathbf{z}_{t+1}$ をサンプリング

$$\mathbf{z}_{t+1}^{(k)} \sim p(\mathbf{z}_{t+1} \mid \mathbf{z}_t^{(l_k)}) \qquad k = 1, \ldots, L \text{ まで繰り返す}$$

SIR フィルタの流れ



リサンプリングの1実装

Systematic Resampling

– 混合分布の重みからのサンプリング方法
– 重みに比例して規則的にサンプリング
  • サンプル数 L のもとで

1. $\hat{p}(\mathbf{z}_t \mid \mathbf{x}_{1:t})$ の累積分布を計算

$$c_1 = 0$$
$$c_l = c_{l-1} + w_t^{(l)} \qquad (l = 1, \ldots, L)$$

2. 一様乱数を引く　$u_1 \sim \mathrm{U}[0, L^{-1}]$

リサンプリングの1実装

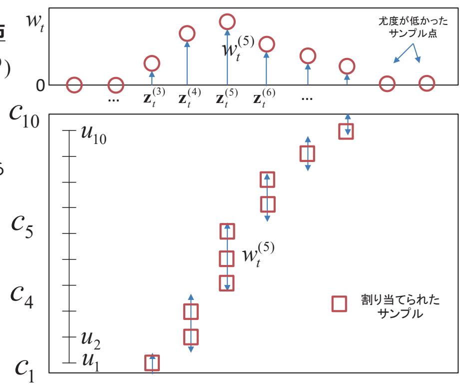
3. サンプルの割り当て → 重みをリセット

例: $L = 10$ の場合の事後分布

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \sum_{l=1}^{L} w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)})$$

乱数で決めたオフセットから
等間隔でサンプリング

$$u_j = u_1 + L^{-1}(j-1)$$
$$(j = 1, \ldots, L)$$



SIRアルゴリズムのまとめ

時刻 t 毎に下記を繰り返し

1. $\mathbf{z}_{t-1}^{(l)}$ に基づきサンプリング $(l = 1, \ldots, L)$

$$\mathbf{z}_t^{(l)} \sim p(\mathbf{z}_t \mid \mathbf{z}_{t-1}^{(l)})$$ もしくは $$\mathbf{w}_{t-1}^{(l)} \sim p(\mathbf{w}_{t-1})$$ $$\mathbf{z}_t^{(l)} = \mathbf{f}_{t-1}(\mathbf{z}_{t-1}^{(l)}, \mathbf{w}_{t-1}^{(l)})$$

※ 乱数生成＋モデルよる予測

2. 尤度で重み計算

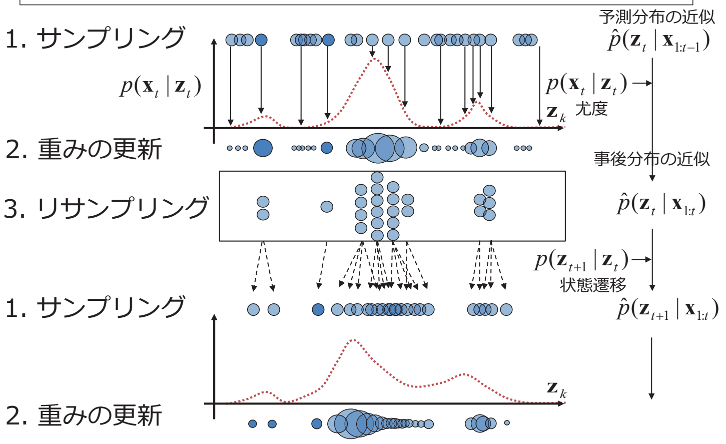
$$w_t^{(l)} = \frac{p(\mathbf{x}_t \mid \mathbf{z}_t^{(l)})}{\sum_{m=1}^{L} p(\mathbf{x}_t \mid \mathbf{z}_t^{(m)})}$$

3. $\mathbf{z}_t^{(l')}$ をリサンプリング $(l' = 1, \ldots, L)$

$$\mathbf{z}_t^{(l')} \sim \sum_{l=1}^{L} w_t^{(l)} \delta(\mathbf{z}_t - \mathbf{z}_t^{(l)})$$

更新後, 添え字を付けなおす
$(l' \to l)$

SIR フィルタの流れ（再掲）



応用する場合に考えてみるべき点

状態ベクトルの次元数

– 巨大な場合: 計算量・粒子数
– 観測ベクトルの次元数より大きい場合

モデル化

– センサデータそのまま or 特徴量表現
  • 低次元潜在空間の導入といった選択肢もありえる
– 人の知識・知見を取り込める部分はどこか
  • 物理モデル，状態変数の分解（問題に合わせた確率的生成モデルの仮定），事前分布
– データから学習させないと難しい部分はどこか
– 不確定要素を含む部分はどこか

次回以降
モデルパラメータの学習を含む

今回の話: パラメータが既知 $\mathbf{A}, \mathbf{C}$ や $p(\mathbf{x}_t \mid \mathbf{z}_t)$ $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$

– **"データ"と"パラメータ"の関係**

- モデル: $\boldsymbol{y} = \boldsymbol{f}(\mathbf{x}, \boldsymbol{\Theta}), p(\boldsymbol{y}|\mathbf{x}, \boldsymbol{\Theta})$
- データ $\mathbf{x}$ とパラメータ $\boldsymbol{\Theta}$ 固定で $\boldsymbol{y}$ を求める
  → 認識・予測・推論
- データ $(\boldsymbol{y}, \mathbf{x})$ 固定でパラメータ $\boldsymbol{\Theta}$ を求める → 学習

次回以降: データを用いて決定 = 学習

– 意識すべき 3つの要素

モデル × コスト関数 × 学習（最適化）

型　「良さ」の定義　値の求め方

参考文献

- C.M.ビショップ「パターン認識と機械学習 上下」
- 持橋大地，大羽成征「ガウス過程と機械学習」
- 浅野太「音のアレイ信号処理」
- S. ヘイキン（武部 幹訳）「適応フィルタ入門」
- S. Haykin, "Adaptive Filter Theory fifth edition"

参考文献

密度関数推定 using NNs

D. A. Nix, A.S. Weigend, "Learning Local Error Bars for Nonlinear Regression," *in Advances Neural Information Processing Systems 7*, 1994.

P. M. Williams, "Using Neural Networks to Model Conditional Multivariate Densities," Neural Computation, vol. 8, no. 4, pp. 843-854, 1996.

状態空間モデル using NNs

R. G. Krishnan, U. Shalit, D. Sontag, "Structured Inference Networks for Nonlinear State Space Models," *in Proc. of AAAI*, 2017, pp.2101-2109.

G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. van Sloun and Y. C. Eldar, "KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics," in IEEE Transactions on Signal Processing, vol. 70, pp. 1532-1547, 2022.

Unscented Particle Filter

R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan, "The Unscented Particle Filter," *in Advances Neural Information Processing Systems 13*, 2000.