

テーマ7: 確率的モデル化と統計的推論

1 目的

例題を通して、確率モデル、統計的推論、動的計画法、事後確率最大化など、パターン認識に重要な役割を果たす手法についての理解を深める。

キーワード:

正規分布, マルコフ連鎖, 乱数, ベイズ推定, 事後確率最大化, 隠れマルコフモデル, 動的計画法

2 連絡事項

- 担当教員: 伊達 章 (A-423, date@cs.miyazaki-u.ac.jp, 内線 7986), 片山 晋
- TA: 岩岡 広大
- 特別演習テーマ7「確率的モデル化と統計的推論」の web ページ:
<http://www.cs.miyazaki-u.ac.jp/~date/lectures/2007tokuen/index.html>
サンプルプログラム, データなどはこのページに置きます。
- 質問は随時受け付けます (出張などで不在の時もあるので, あらかじめメールで時間を予約するほうが確実)。

成績は, レポートおよび口頭説明 (それぞれ2回づつ)と, 最終日におこなう 自由課題の発表により評価。内容が理解できているかどうかを重視。

- この特別演習課題には, ある一つの山がある。その山は, 話を聞いただけで乗り越えられる (理解できる) 類のものではない。理解できたかどうかは, 自分でプログラムを書いて, それを走らせることにより, 確認できる。うまくいった場合, 理解が深まった! と実感できるはずである。ここまでできれば合格で, そこからは自由に課題を考えればいい。理解しているかどうかは一目瞭然なので, 他人のプログラムをコピーするなどのごまかしはいっさい効かないが, 理解できないところがあれば, 友達に聞いて理解するのはいい。
- プログラムにバグがあり, どこに問題があるか発見するまで相当の時間がかかることがある。この作業は忍耐を要する。担当教員, TA もできる限りバグ探しを助けるが, まずは時間外に相当の時間をかけて独力で探すこと。
- 分からなければ, 締め切り間際でなく, 早めに相談に来ること。
- 課題が簡単すぎると感じた場合も, 適宜それなりの課題を与えるので, 相談に来ること。
- このテーマをよりよいものにしていきたいので, 改善案がある人は, すぐに担当教員に伝えること。

3 日程

4月11日 問題の理解，コンピュータ環境への適応，基礎的事項の確認

データの生成をおこなうプログラムを書くことにより，一様分布，正規分布，マルコフ性，条件付確率の理解を確認する．休み明けなので，エディタの使い方，C言語，プログラムのコンパイルの仕方，gnuplotの使い方を思い出す．その他，同時分布，周辺分布など確率・統計の基本概念を復習する．

- 本日の到達目標：図2，図4に示すようなデータを各自で生成し表示する．
 1. 計算機で（疑似）乱数を生成する．[]
 2. マルコフ的情報源を計算機で生成する．図2のようなデータを生成．[]
 3. 計算機で正規分布にしたがうデータを生成する．[]
 4. マルコフ的情報源にノイズを加える．図4のようなデータを生成．[]
 5. 確率・統計の復習：同時確率分布，周辺確率分布，ベイズの定理 []
- 配布資料
 - * この資料：もとの信号．ノイズを含んだ図．復元した図．
 - * 速修 正規分布，正規分布生成のコード例
 - * 速修 確率・統計
 - * gnuplot グラフ生成のスクリプト

課題1（レポートと口頭．締切：4月18日）：

標準正規分布に従う大量のデータを計算機で出力し，そのデータが正規分布にしたがっていることをそれぞれの分布の性質を利用し確認する（課題1-1）．計算機で実際にどう乱数を発生させているか，発生する方法があるか，調べてみる（課題1-2）．基本課題について，どのような方法でもいいので，元データの復元を試みる（課題1-3）．課題1-1は必須．

● 4月18日 知識習得とコード作成

- 知識習得：ベイズ推定．動的計画法による事後確率を最大化する推定値の求め方．
- プログラムコード作成．

課題2（レポートと口頭．締切：5月2日の演習が始まるまで）：

基本事項（事後確率の計算の仕方，動的計画法を理解し，個別に与えられたデータに対し，問題を解決するコードを書く．

● 4月25日 基本課題のコード作成

● 5月2日 課題3（自由課題）モデル化：各自で面白い問題を考え，コードを作成する．

● 5月9日 自由課題のコード作成，分析，発表準備

● 5月16日 発表準備

● 5月23日 自由課題の成果発表：発表で使った資料を5月30日までに提出．

4 例

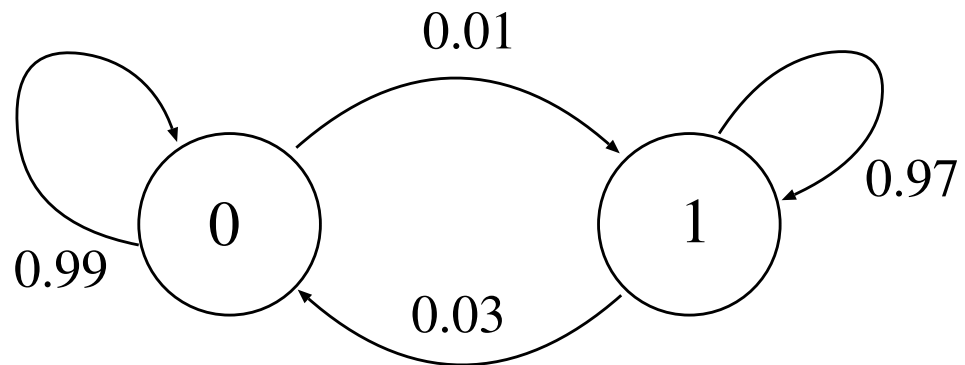


図 1: マルコフの情報源 . 状態遷移図

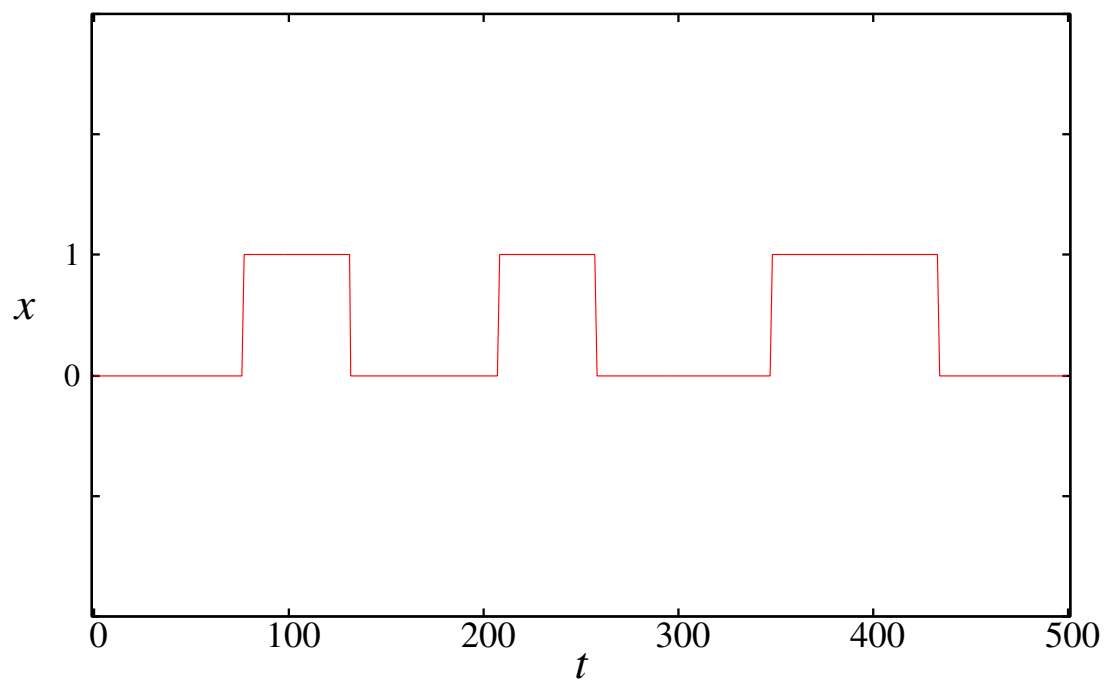


図 2: ランダムサンプル

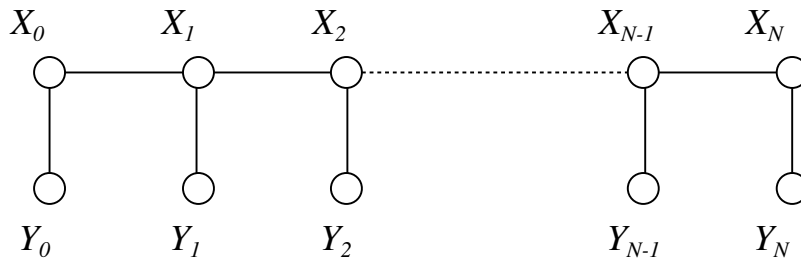


図 3: 確率変数の依存性を描いたグラフ

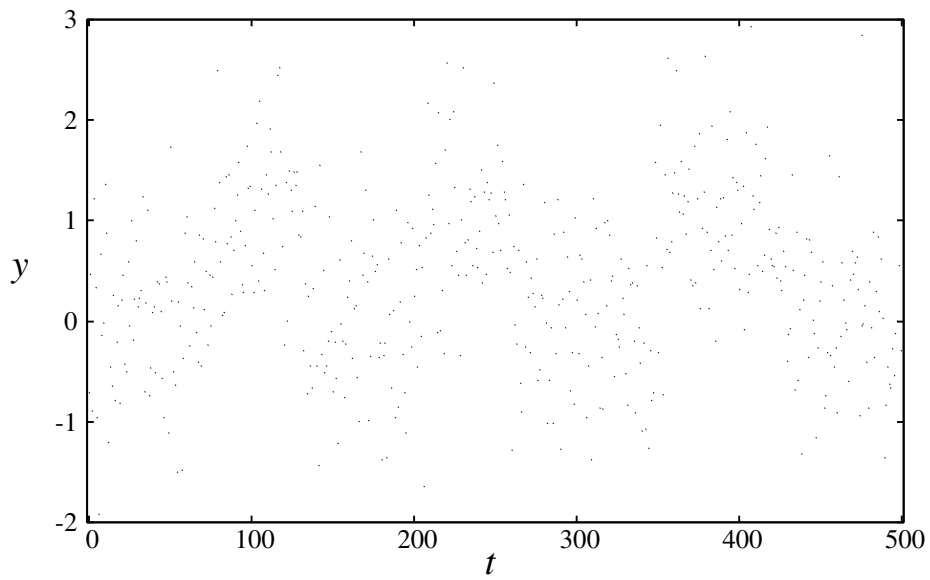


図 4: 観測値

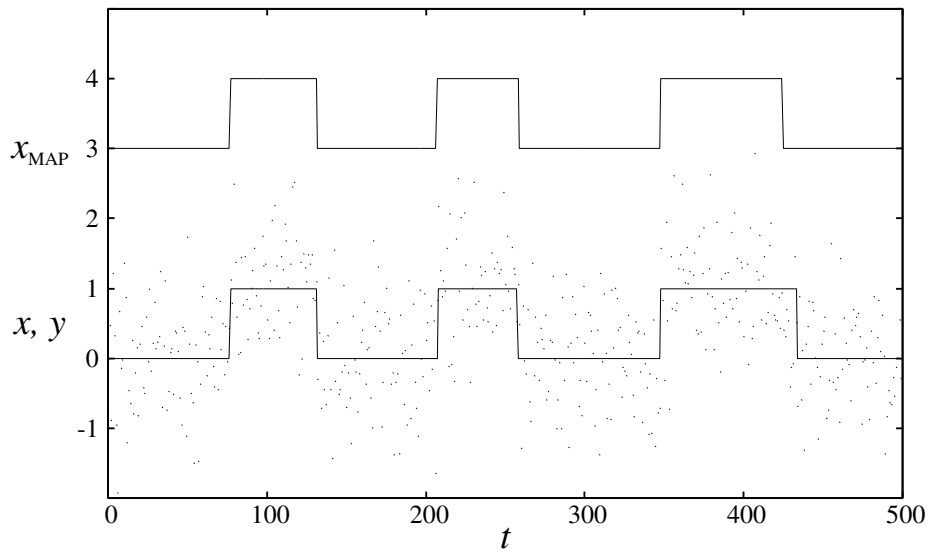


図 5: 事後確率最大にする値 x_{MAP}