九州大学 工学部地球環境工学科
船舶海洋システム工学コース

システム設計工学 （担当：木村）

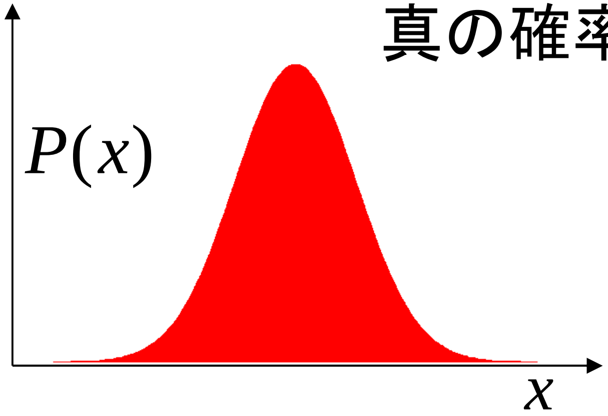
（12） データからのマルコフモデル構築

場所：船1講義室

http://sysplan.nams.kyushu-u.ac.jp/gen/index.html

【復習】確率モデルの推定

環境



真の確率モデル（母関数） 例：正規分布

$$P(x) = N(\mu, \sigma^2)$$

母平均（期待値）および母分散などのパラメータは観測者にとって未知

観測者は真のモデルが正規分布であることは知っている

観測データ
標本(sample) $x_1, x_2, \cdots, x_n$

観測者

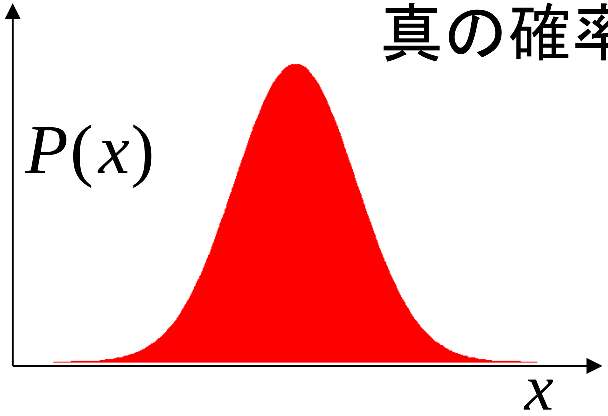
標本を用いて、未知パラメータを推定する： 推定値 $\hat{\mu}, \hat{\sigma}$

単一のパラメータ推定値を求める →

未知パラメータが存在する区間を求める →

【復習】確率モデルの推定

環境



真の確率モデル（母関数） 例：正規分布

$$P(x) = N(\mu, \sigma^2)$$

母平均（期待値）および母分散などの
パラメータは観測者にとって未知

観測者は真のモデルが
正規分布であることは
知っている

観測データ
標本(sample) $x_1, x_2, \cdots, x_n$

観測者

標本を用いて、未知パラメータを推定する： 推定値 $\hat{\mu}, \hat{\sigma}$

単一のパラメータ推定値を求める → **点推定**（最尤推定法など）

未知パラメータが存在する区間を求める → **区間推定**

ゆう

最尤推定法(maximum likelihood method)とは？

推定モデル $\hat{P}(x)$ から観測データ $x_1, x_2, \cdots, x_n$ が生成される確率 が最大となるように推定パラメータ θ を決定する

モデルの「もっともらしさ」を表す量

例） $P(x)$が正規分布なら θ は μ と σ に相当

$$L(\theta) = \hat{P}(x_1)\hat{P}(x_2)\cdots\hat{P}(x_n)$$

この関数を最大化するパラメータを見つければ良いが、**乗算**なので解析しにくい

対数をとる

$$\ln L(\theta) = \ln \hat{P}(x_1) + \ln \hat{P}(x_2) + \cdots + \ln \hat{P}(x_n)$$

線形になっているので、解析しやすい

尤度を最大化する θ と対数尤度を最大化する θ の値は同じになる

対数尤度を最大化するパラメータθを探索するのが一般的

ゆう

最尤推定法(maximum likelihood method)とは？

推定モデル $\hat{P}(x)$ から観測データ $x_1, x_2, \cdots, x_n$ が生成される確率（**尤度**）が最大となるように推定パラメータ θ を決定する

ゆうど

例） P(x)が正規分布なら
θ は μ と σ に相当

モデルの「もっともらしさ」
を表す量

$$L(\theta) = \hat{P}(x_1)\hat{P}(x_2)\cdots\hat{P}(x_n)$$

この関数を最大化するパラメータを見つければ良いが、**乗算**なので解析しにくい

対数をとる

$$\ln L(\theta) = \ln \hat{P}(x_1) + \ln \hat{P}(x_2) + \cdots + \ln \hat{P}(x_n)$$

対数尤度

たいすう　ゆうど

線形になっているので、解析しやすい

尤度を最大化する θ と対数尤度を最大化する θ の値は同じになる

対数尤度を最大化するパラメータθ を探索するのが一般的

例）Bernoulli分布の最尤推定

真の確率分布モデル（母関数）：　確率 p で $x=1$, 確率 $1-p$ で $x=0$ となる分布 $F(x)$

母関数からの標本：　$x_1=0,\ x_2=1,\ x_3=1,\ x_4=0,\ldots,x_n=1$

サンプルnコのうち、$x=0$ となる場合がk回

このとき、未知パラメータpを最尤推定する。このとき**尤度L**は　$L =$

モデルから標本が
発生する確率

対数尤度を計算すると、

$$\ln L(\theta) = \ln F(x_1) + \ln F(x_2) + \cdots + \ln F(x_n)$$
$$= k(\ln(1-p)) + (n-k)(\ln(p))$$

これを最大化するpを求めるためpで微分

$$\ln F(x) = \begin{cases} \ln(1-p) & \text{where} \quad x=0 \\ \ln(\mathrm{p}) & \text{where} \quad x=1 \end{cases}$$

$$\frac{\partial}{\partial p}\ln F(x) = \begin{cases} \dfrac{-1}{1-p} & \text{where} \quad x=0 \\ \dfrac{1}{p} & \text{where} \quad x=1 \end{cases}$$

$$\frac{\partial}{\partial p}\ln L(\theta) =$$

最大点では傾きゼロ
これを解くと、
　p=（$x=1$となる場合の数）／（全標本数）
という常識的な式を得る

例）Bernoulli分布の最尤推定

真の確率分布モデル（母関数）：　確率 p で $x=1$, 確率 $1-p$ で $x=0$ となる分布 $F(x)$

母関数からの標本：　$x_1=0, x_2=1, x_3=1, x_4=0, \ldots, x_n=1$
サンプルnコのうち、$x=0$ となる場合がk回

このとき、未知パラメータpを最尤推定する。このとき**尤度L**は　$L = \boxed{(1-p)^k \; p^{n-k}}$

モデルから標本が
発生する確率

対数尤度を計算すると、

$$\ln L(\theta) = \ln F(x_1) + \ln F(x_2) + \cdots + \ln F(x_n)$$
$$= k(\ln(1-p)) + (n-k)(\ln(p))$$

これを最大化するpを求めるためpで微分

$$\frac{\partial}{\partial p}\ln L(\theta) = \boxed{k\left(\frac{-1}{1-p}\right) + (n-k)\frac{1}{p}}$$

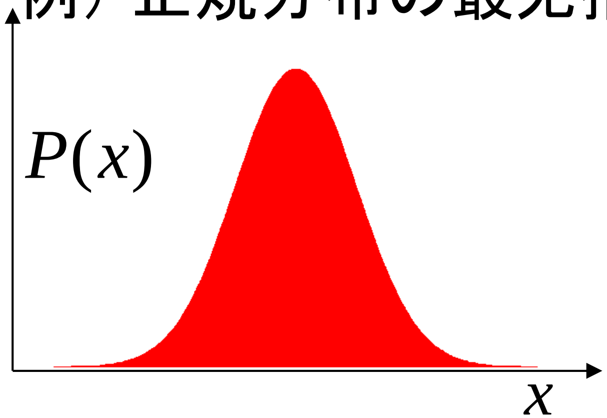
$$\ln F(x) = \begin{cases} \ln(1-p) & \text{where} \quad x=0 \\ \ln(\mathrm{p}) & \text{where} \quad x=1 \end{cases}$$

$$\frac{\partial}{\partial p}\ln F(x) = \begin{cases} \dfrac{-1}{1-p} & \text{where} \quad x=0 \\ \dfrac{1}{p} & \text{where} \quad x=1 \end{cases}$$

$$\boxed{\begin{aligned} & k\left(\frac{-1}{1-p}\right) + (n-k)\frac{1}{p} = 0 \\ & p = \frac{n-k}{n} \end{aligned}}$$

最大点では傾きゼロ
これを解くと、
　p=（$x=1$となる場合の数）／（全標本数）
という常識的な式を得る

例）正規分布の最尤推定



真の確率モデル：正規分布

$$P(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

標本を用いて、未知パラメータ $\hat{\mu}, \hat{\sigma}$ を最尤推定した場合、

右式のように標本平均と標本分散より
推定値が得られることを示せ。

$$\begin{cases} \hat{\mu} = \dfrac{x_1 + x_2 + \cdots x_n}{n}, \\ \hat{\sigma}^2 = \dfrac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_n - \hat{\mu})^2}{n} \end{cases}$$

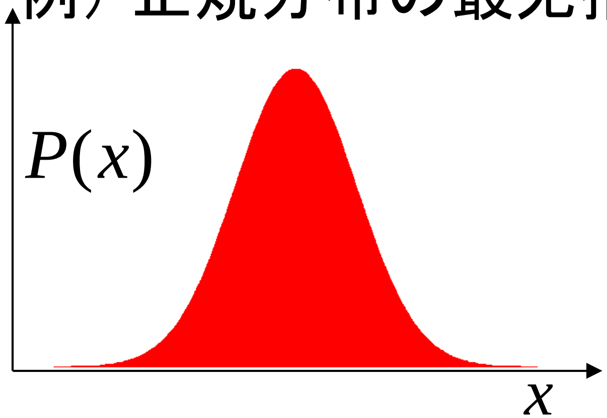
（ヒント） 微分公式

$$\frac{\partial}{\partial \theta} \ln f(\theta) = \frac{1}{f(\theta)} \frac{\partial}{\partial \theta} f(\theta)$$

確率密度関数の場合は、密度関数値＝尤度で計算する：

モデル *f(x)* から
標本が発生する確率

例）正規分布の最尤推定



真の確率モデル:正規分布

$$f(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

標本を用いて、未知パラメータ $\hat{\mu}, \hat{\sigma}$ を最尤推定した場合、

右式のように標本平均と標本分散より推定値が得られることを示せ。

$$\begin{cases} \hat{\mu} = \dfrac{x_1 + x_2 + \cdots x_n}{n}, \\ \hat{\sigma}^2 = \dfrac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_n - \hat{\mu})^2}{n} \end{cases}$$

（ヒント） 微分公式

$$\frac{\partial}{\partial\theta} \ln f(\theta) = \frac{1}{f(\theta)} \frac{\partial}{\partial\theta} f(\theta)$$

確率密度関数の場合は、密度関数値＝尤度で計算する:

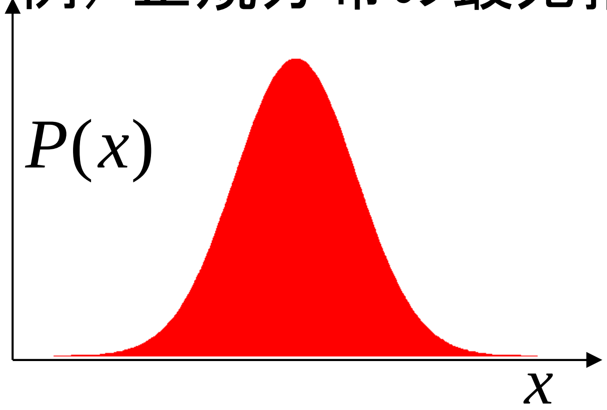
モデル *f(x)* から
標本が発生する確率

$$\ln L(\mu, \sigma) = \ln f(x_1) + \ln f(x_2) + \cdots + \ln f(x_n)$$

$$= \sum_{i=1}^{n} \ln f(x_i) = \sum_{i=1}^{n} \ln \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right)$$

$$\frac{\partial}{\partial\mu} \ln L(\mu, \sigma) = \sum_{i=1}^{n} -\frac{(x_i - \mu)}{\sigma^2} = 0$$ を解くと、

例）正規分布の最尤推定



真の確率モデル：正規分布

$$f(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

これから得られる観測データ

標本(sample)　$x_1, x_2, \cdots, x_n$　のとき、

標本を用いて、未知パラメータ $\hat{\mu}, \hat{\sigma}$ を最尤推定した場合、

右式のように標本平均と標本分散より推定値が得られることを示せ。

$$\begin{cases} \hat{\mu} = \dfrac{x_1 + x_2 + \cdots x_n}{n}, \\ \hat{\sigma}^2 = \dfrac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_n - \hat{\mu})^2}{n} \end{cases}$$

（ヒント）　微分公式

$$\frac{\partial}{\partial\theta} \ln f(\theta) = \frac{1}{f(\theta)} \frac{\partial}{\partial\theta} f(\theta)$$

確率密度関数の場合は、密度関数値＝尤度で計算する：

$$\ln L(\mu, \sigma) = \ln f(x_1) + \ln f(x_2) + \cdots + \ln f(x_n)$$

$$= \sum_{i=1}^{n} \ln f(x_i) = \sum_{i=1}^{n} \ln \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right)$$

$$\frac{\partial}{\partial\mu} \ln L(\mu, \sigma) = \sum_{i=1}^{n} -\frac{(x_i - \mu)}{\sigma^2} \quad = 0$$

を解くと、

$$\sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\sum_{i=1}^{n} x_i = n\mu$$

$$\frac{\sum_{i=1}^{n} x_i}{n} = \mu$$

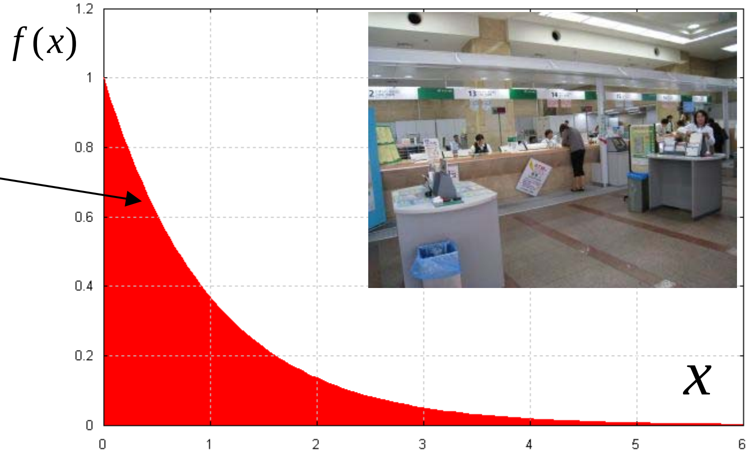
よってμの最尤推定値は
標本平均で与えられる

分散についても同様

【練習問題】

指数分布の
確率密度関数

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & (x \geq 0)\\ 0 & (x<0)\end{cases}$$



これから得られる観測データ
標本(sample) $x_1, x_2, \cdots, x_n$ のとき、
パラメータλを最尤推定せよ。

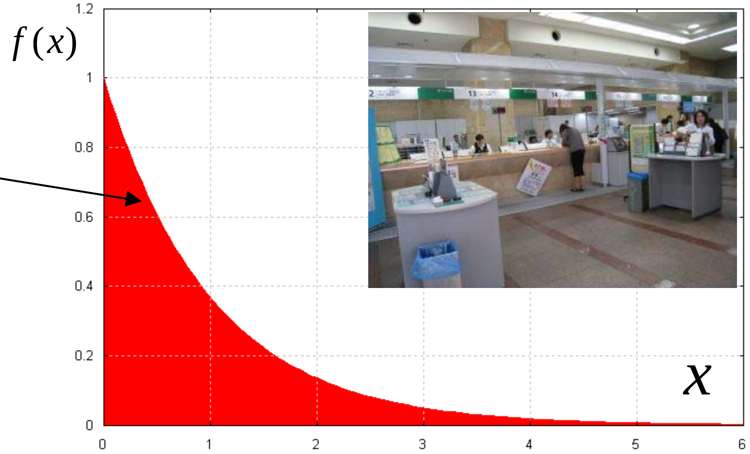
確率密度関数の場合は、密度関数値＝尤度で計算する:

モデル *f(x)* から
標本が発生する確率

【練習問題】

指数分布の
確率密度関数

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & (x\geq 0)\\ 0 & (x<0)\end{cases}$$



これから得られる観測データ
標本(sample) $x_1, x_2, \cdots, x_n$ のとき、
パラメータλを最尤推定せよ。

確率密度関数の場合は、密度関数値＝尤度で計算する:

モデル *f(x)* から
標本が発生する確率

$$\ln L(\lambda)=\ln f(x_1)+\ln f(x_2)+\cdots+\ln f(x_n)$$

$$=\sum_{i=1}^{n}\ln f(x_i)=\sum_{i=1}^{n}\ln\lambda\exp(-\lambda x_i)$$

$$\frac{\partial}{\partial\lambda}\ln L(\lambda)=\sum_{i=1}^{n}\frac{1}{\lambda\exp(-\lambda x_i)}\left(\exp(-\lambda x_i)+\lambda\exp(-\lambda x_i)(-x_i)\right)$$

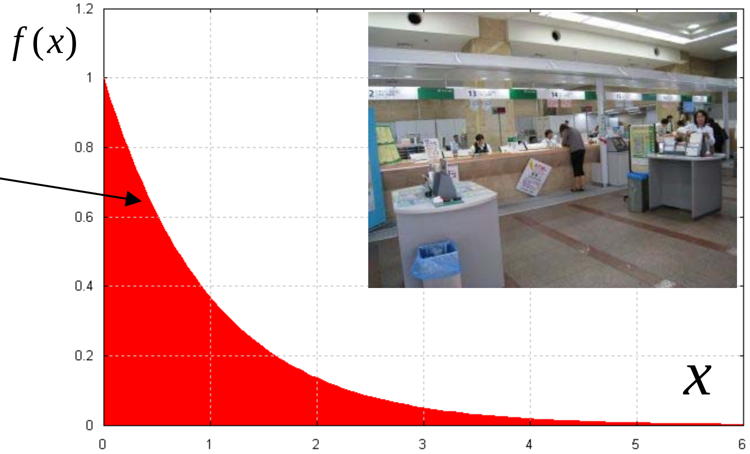
$$=\sum_{i=1}^{n}\frac{1}{\lambda}-x_i=0$$

よって

【練習問題】

指数分布の
確率密度関数

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & (x\geq 0)\\ 0 & (x<0)\end{cases}$$



これから得られる観測データ
標本(sample) $x_1, x_2, \cdots, x_n$ のとき、
パラメータλを最尤推定せよ。

確率密度関数の場合は、密度関数値＝尤度で計算する:

モデル *f(x)* から
標本が発生する確率

$$\ln L(\lambda)=\ln f(x_1)+\ln f(x_2)+\cdots+\ln f(x_n)$$

$$=\sum_{i=1}^{n}\ln f(x_i)=\sum_{i=1}^{n}\ln \lambda \exp(-\lambda x_i)$$

$$\frac{\partial}{\partial\lambda}\ln L(\lambda)=\sum_{i=1}^{n}\frac{1}{\lambda\exp(-\lambda x_i)}\left(\exp(-\lambda x_i)+\lambda\exp(-\lambda x_i)(-x_i)\right)$$

$$=\sum_{i=1}^{n}\frac{1}{\lambda}-x_i=0$$

よって $\dfrac{n}{\lambda}-\sum_{i=1}^{n}x_i=0$

$$\frac{n}{\lambda}=\sum_{i=1}^{n}x_i$$

$$\frac{1}{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}$$

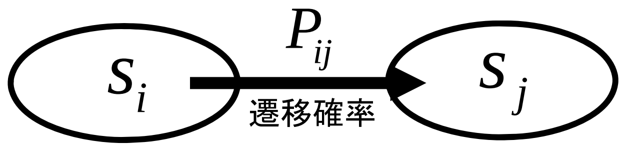
λ :平均到着(故障)率(人／時)

$1/\lambda$:平均到着(故障)時間間隔

よって λ の最尤推定値は
標本平均の逆数で与えられる

マルコフ過程の最尤推定の準備　マルコフモデルからデータが生成される確率

マルコフモデルにおける状態 S_i から S_j への遷移確率を
P_{ij} と表す。



【練習問題0】 状態 S_A から S_B へ遷移する確率 ＝ $\boxed{\quad\quad}$

【練習問題1】 状態 S_A から S_B へ遷移し、その次に S_C へ遷移する確率 ＝ $\boxed{\quad\quad}$

【練習問題1 ‘】 マルコフモデルに従って、状態 S_A からスタートして2回状態遷移したとき、
状態遷移系列 $S_A\ S_B\ S_C$ が観測される確率 ＝ $\boxed{\quad\quad}$

【練習問題2】 状態が3つ（ S_A, S_B, S_C ）のマルコフモデルに従って、ある決まった状態から
スタートして 20 回状態遷移したとき、

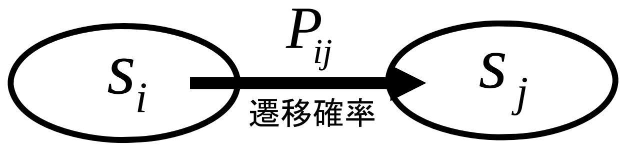
s_A から s_A への遷移が　2回
s_A から s_B への遷移が　2回
s_A から s_C への遷移が　3回
s_C から s_A への遷移が　3回
s_C から s_B への遷移が　3回
s_C から s_C への遷移が　2回

s_B から s_A への遷移が　1回
s_B から s_B への遷移が　1回
s_B から s_C への遷移が　3回

このような遷移が観測される確率（尤度）
＝ $\boxed{\quad\quad}$

マルコフ過程の最尤推定の準備　マルコフモデルからデータが生成される確率

マルコフモデルにおける状態 S_i から S_j への遷移確率を P_{ij} と表す。



【練習問題0】 状態 S_A から S_B へ遷移する確率 ＝ $\boxed{P_{AB}}$

【練習問題1】 状態 S_A から S_B へ遷移し、その次に S_C へ遷移する確率 ＝ $\boxed{P_{AB}P_{BC}}$

【練習問題1 ‘】 マルコフモデルに従って、状態 S_A からスタートして2回状態遷移したとき、状態遷移系列 $S_A\ S_B\ S_C$ が観測される確率 ＝ $\boxed{P_{AB}P_{BC}}$

【練習問題2】 状態が3つ（ S_A, S_B, S_C ）のマルコフモデルに従って、ある決まった状態からスタートして 20 回状態遷移したとき、

s_A から s_A への遷移が　2回
s_A から s_B への遷移が　2回
s_A から s_C への遷移が　3回
s_C から s_A への遷移が　3回
s_C から s_B への遷移が　3回
s_C から s_C への遷移が　2回

s_B から s_A への遷移が　1回
s_B から s_B への遷移が　1回
s_B から s_C への遷移が　3回

このような遷移が観測される確率（尤度）

＝ $\boxed{P_{AA}^2 P_{AB}^2 P_{AC}^3 P_{BA}^1 P_{BB}^1 P_{BC}^3 P_{CA}^3 P_{CB}^3 P_{CC}^2}$

マルコフ過程の最尤推定

観測データからマルコフモデルを構築する

観測される状態遷移の履歴を $x_0 x_1 \cdots x_T$ と表し、状態 i から j への推移回数を n_{ij} と表す。
マルコフモデルにおける状態 i から j へ遷移確率を P_{ij} と表す。
すると観測された履歴が発生する確率（尤度）は

$$L = P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{T-1} x_T} = \prod_{i=1}^{S} \prod_{j=1}^{S} P_{ij}^{\ n_{ij}}$$

で与えられる。これを最大にする P_{ij} を求める

確率なので
値は非負

ただし、Pij は条件として

$$\begin{cases} P_{ij} \geq 0, \quad i, j = 1,2,\cdots,S \\ \sum_{j=1}^{S} Pij = 1, \quad i, j = 1,2,\cdots,S \end{cases}$$

を満たしていなければならない

確率を合計
すると1になる

等式を含む $\boxed{\text{制約条件付き最大値問題}}$

↓

$\boxed{}$ を使って解く

マルコフ過程の最尤推定

観測データからマルコフモデルを構築する

観測される状態遷移の履歴を $x_0 x_1 \cdots x_T$ と表し、状態 i から j への推移回数を n_{ij} と表す。
マルコフモデルにおける状態 i から j へ遷移確率を P_{ij} と表す。
すると観測された履歴が発生する確率（尤度）は

$$L = P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{T-1} x_T} = \prod_{i=1}^{S} \prod_{j=1}^{S} P_{ij}^{\ n_{ij}}$$

で与えられる。これを最大にする P_{ij} を求める

確率なので
値は非負

ただし、Pij は条件として

$$P_{ij} \geq 0, \quad i, j = 1, 2, \cdots, S$$

$$\sum_{j=1}^{S} Pij = 1, \quad i, j = 1, 2, \cdots, S$$

を満たしていなければならない

確率を合計
すると1になる

等式を含む **制約条件付き最大値問題**

ラグランジュ未定乗数法 を使って解く

（参考）ラグランジュ未定乗数法 とは？

N 個の変数 $\mathbf{X}=[\theta_1,\theta_2,\cdots\theta_N]$ について、M 個の制約条件 $g_i(\mathbf{X})=0 \quad (i=1,2,\cdots M)$
が与えられている。
この制約のもとで、ある関数 $f(\mathbf{X})$ が極値をとるような変数 $\mathbf{X}$ を求める。

M 個の別の未知変数 $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ を使って、以下の関数 $F(\mathbf{X},\boldsymbol{\lambda})$ を考える:

$$F(\mathbf{X},\boldsymbol{\lambda})=f(\mathbf{X})+\sum_{i=1}^{M}\lambda_i\, g_i(\mathbf{X}) \qquad \text{（式1）}$$

この関数 $F(\mathbf{X},\boldsymbol{\lambda})$ の極値条件は、

$$\frac{\partial}{\partial\theta_k}F(\mathbf{X},\boldsymbol{\lambda})=\frac{\partial}{\partial\theta_k}f(\mathbf{X})+\frac{\partial}{\partial\theta_k}\sum_{i=1}^{M}\lambda_i\, g_i(\mathbf{X})=0 \qquad (k=1,2,\cdots N)$$

$$\frac{\partial}{\partial\lambda_i}F=g_i(\mathbf{X})=0 \qquad (i=1,2,\cdots M)$$

これらを満たす解の中に、求める解が存在する。

ここで式（１）をラグランジュ関数、$\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ をラグランジュの未定乗数という。

（参考）ラグランジュ未定乗数法とは？

N 個の変数 $\mathbf{X}=[\theta_1,\theta_2,\cdots\theta_N]$ について、M 個の制約条件 $g_i(\mathbf{X})=0 \quad (i=1,2,\cdots M)$
が与えられている。
この制約のもとで、ある関数 $f(\mathbf{X})$ が極値をとるような変数 $\mathbf{X}$ を求める。

M 個の別の未知変数 $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ を使って、以下の関数 $F(\mathbf{X},\boldsymbol{\lambda})$ を考える:

$$F(\mathbf{X},\boldsymbol{\lambda}) = \underbrace{f(\mathbf{X})}_{\text{最大化・最小化したい値}} + \sum_{i=1}^{M} \lambda_i \underbrace{g_i(\mathbf{X})}_{\text{制約条件}} \qquad \text{（式１）}$$

この関数 $F(\mathbf{X},\boldsymbol{\lambda})$ の極値条件は、

$$\frac{\partial}{\partial\theta_k}F(\mathbf{X},\boldsymbol{\lambda}) = \frac{\partial}{\partial\theta_k}f(\mathbf{X}) + \frac{\partial}{\partial\theta_k}\sum_{i=1}^{M}\lambda_i\, g_i(\mathbf{X}) = 0 \qquad (k=1,2,\cdots N)$$

$$\frac{\partial}{\partial\lambda_i}F = g_i(\mathbf{X}) = 0 \qquad (i=1,2,\cdots M)$$

これらを満たす解の中に、求める解が存在する。

ここで式（１）をラグランジュ関数、$\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ をラグランジュの未定乗数という。

マルコフ過程の最尤推定

ラグランジュ関数は

$$F = \underline{\log L} + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{j=1}^{S} P_{ij} \right) - 1 \right)$$

尤度 $L = P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{T-1} x_T} = \prod_{i=1}^{S} \prod_{j=1}^{S} P_{ij}^{\,n_{ij}}$

代入

より、対数尤度 $\log L = \boxed{\quad}$

まだ未知数 λ_i を含んでいる

$\log L$ を最大にする P_{ij} は、$\boxed{\quad}$ を $\boxed{\quad}$ と $\boxed{\quad}$ で偏微分し、それぞれをゼロと置いた連立方程式を解けば良い。

$$\frac{\partial F}{\partial P_{ij}} = \boxed{\quad}$$

この式の右辺を0とおくと、$P_{ij} = \boxed{\quad}$ が得られる。

$$\frac{\partial F}{\partial \lambda_i} = \boxed{\quad}$$

この式の右辺を0とおき、上の式と連立して解くと

$$\lambda_i = \boxed{\quad}$$

λ_i を消去すると、

$$P_{ij} = \boxed{\quad}$$

この解は非負条件を満たしている

マルコフ過程の最尤推定

ラグランジュ関数は

$$F = \underline{\log L} + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{j=1}^{S} P_{ij} \right) - 1 \right)$$

代入

尤度　$L = P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{T-1} x_T} = \prod_{i=1}^{S} \prod_{j=1}^{S} P_{ij}^{\,n_{ij}}$

より、対数尤度　$\log L = \boxed{\sum_{i=1}^{S} \sum_{j=1}^{S} n_{ij} \log P_{ij}}$

まだ未知数 λ_i を含んでいる

$\log L$ を最大にする P_{ij} は、$\boxed{F}$ を $\boxed{P_{ij}}$ と $\boxed{\lambda_i}$ で偏微分し、それぞれをゼロと置いた連立方程式を解けば良い。

$$\frac{\partial F}{\partial P_{ij}} = \boxed{\frac{n_{ij}}{P_{ij}} + \lambda_i}$$

この式の右辺を0とおくと、$P_{ij} = \boxed{-\dfrac{n_{ij}}{\lambda_i}}$ が得られる。

$$\frac{\partial F}{\partial \lambda_i} = \boxed{\left(\sum_{j=1}^{S} P_{ij} \right) - 1}$$

この式の右辺を0とおき、上の式と連立して解くと

$$\lambda_i = \boxed{-\sum_{j=1}^{S} n_{ij}}$$

λ_i を消去すると、

$$P_{ij} = \boxed{\frac{n_{ij}}{\sum_{k=1}^{S} n_{ik}}}$$

この解は非負条件を満たしている

マルコフ過程の最尤推定（結論）

観測される状態遷移の履歴を $x_0 x_1 \cdots x_T$ と表し、状態 i から j への推移回数を n_{ij} と表す。
マルコフモデルにおける状態 i から j へ遷移確率を P_{ij} と表す。

このときデータから推定される P_{ij} の最尤推定値 $\hat{P}_{ij}$ は、

ラグランジュ関数

$$F = \log L + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{j=1}^{S} P_{ij} \right) - 1 \right)$$

対数尤度

と置いて、極値条件から連立方程式を解くと、最尤推定値

$$\hat{P}_{ij} = \boxed{\quad}$$

状態 i からj へ遷移した回数

状態 i を訪問した回数

を得る。

モデルに何らかの制約や構造を含む場合、この式が成り立つとは限らない。
その場合、最尤推定の基礎に立ち戻って計算しなければならない。 演習問題

マルコフ過程の最尤推定(結論)

観測される状態遷移の履歴を $x_0 x_1 \cdots x_T$ と表し、状態 i から j への推移回数を n_{ij} と表す。
マルコフモデルにおける状態 i から j へ遷移確率を P_{ij} と表す。

このときデータから推定される P_{ij} の最尤推定値 $\hat{P}_{ij}$ は、

ラグランジュ関数 $$F = \log L + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{j=1}^{S} P_{ij} \right) - 1 \right)$$

対数尤度

と置いて、極値条件から連立方程式を解くと、最尤推定値

$$\hat{P}_{ij} = \boxed{\frac{n_{ij}}{\sum_{k=1}^{S} n_{ik}}}$$

状態 i からj へ遷移した回数

状態 i を訪問した回数

を得る。

モデルに何らかの制約や構造を含む場合、この式が成り立つとは限らない。
その場合、最尤推定の基礎に立ち戻って計算しなければならない。 演習問題

【演習問題】2019.01.25

学籍番号

氏　名 ______________________________

２状態のマルコフ過程モデルを考える。

観測データより得られた状態 i から j への遷移回数を n_{ij} と表す。

マルコフモデルにおける状態 i から j への遷移確率を P_{ij} と表す。

ここで、構造上 $P_{11} = P_{22}$ であることが分かっているものとする。

このとき、P_{ij} の最尤推定値を、遷移回数 $n_{11}, n_{12}, n_{21}, n_{22}$ を用いて表せ。

【演習問題】

学籍番号
氏　名

２状態のマルコフ過程モデルを考える。

観測データより得られた状態 i から j への遷移回数を n_{ij} と表す。

マルコフモデルにおける状態 i から j への遷移確率を P_{ij} と表す。

ここで、構造上 $P_{11} = P_{22}$ であることが分かっているものとする。

このとき、P_{ij} の最尤推定値を、遷移回数 $n_{11}, n_{12}, n_{21}, n_{22}$ を用いて表せ。

遷移確率行列は、$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{11} \end{bmatrix}$ ここで $\begin{cases} P_{11} + P_{12} = 1 \\ P_{21} + P_{11} = 1 \end{cases}$ より、

$$P_{12} = P_{21} = 1 - P_{11}$$

したがって

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11} & 1 - P_{11} \\ 1 - P_{11} & P_{11} \end{bmatrix}$$

推定すべき未知変数は
ただ1つ

【演習問題】

学籍番号
氏　名

２状態のマルコフ過程モデルを考える。

観測データより得られた状態 i から j への遷移回数を n_{ij} と表す。

マルコフモデルにおける状態 i から j への遷移確率を P_{ij} と表す。

ここで、構造上 $P_{11} = P_{22}$ であることが分かっているものとする。

このとき、P_{ij} の最尤推定値を、遷移回数 $n_{11}, n_{12}, n_{21}, n_{22}$ を用いて表せ。

遷移確率行列は、$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{11} \end{bmatrix}$ ここで $\begin{cases} P_{11} + P_{12} = 1 \\ P_{21} + P_{11} = 1 \end{cases}$ より、

$$P_{12} = P_{21} = 1 - P_{11}$$

したがって

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11} & 1-P_{11} \\ 1-P_{11} & P_{11} \end{bmatrix}$$

推定すべき未知変数は
ただ1つ

対数尤度

$$\log L = \sum_{i=1}^{S} \sum_{j=1}^{S} n_{ij} \log P_{ij}$$

これを最大化する P_{11} を求める

$$= n_{11} \log P_{11} + n_{12} \log(1 - P_{11}) + n_{21} \log(1 - P_{11}) + n_{22} \log P_{11}$$

$\dfrac{\partial \log L}{\partial P_{11}} = 0$ を計算して整理すると、最尤推定値は $P_{11} = \dfrac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}}$