統計的機械学習
第三回　レポート

37-196360　森田涼介

2019 年 5 月 6 日

宿題 1

微分と積分の順序が交換できるとき，不偏推定量 $\hat{\theta}$ について，次式を示す。

$$\mathbb{E}_{\boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_n \sim q(\cdot;\ \theta)}\left[\hat{\theta} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log q(\boldsymbol{x}_i;\ \theta)\bigg|_{\theta=\theta^*}\right] = 1 \tag{1}$$

i.i.d. な訓練標本 $\{\boldsymbol{x}_i\}_{i=1}^{n}$ がモデル $q(\boldsymbol{x};\ \theta)$ から生起するとき，その確率は，

$$p(\boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_n) = \prod_{i=1}^{n} q(\boldsymbol{x}_i;\ \theta) \tag{2}$$

となる。このとき，対数尤度とその θ による微分は，

$$L(\theta) = \log p = \sum_{i=1}^{n} \log q(\boldsymbol{x}_i;\ \theta) \tag{3}$$

$$\frac{\partial L(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}(\log p) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta}(\log q(\boldsymbol{x}_i;\ \theta)) \tag{4}$$

である。

いま，式 (1) の左辺は次のように変形できる。

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_n \sim q(\cdot;\ \theta)}\left[\hat{\theta} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log q(\boldsymbol{x}_i;\ \theta)\bigg|_{\theta=\theta^*}\right] &= \mathbb{E}_{\boldsymbol{x} \sim p(\cdot;\ \theta)}\left[\hat{\theta}\ \frac{\partial}{\partial \theta}(\log p)\bigg|_{\theta=\theta^*}\right] && (5) \\
&= \int p\hat{\theta}\ \frac{\partial}{\partial \theta}(\log p)\bigg|_{\theta=\theta^*}\ \mathrm{d}\boldsymbol{x} \\
&= \int \hat{\theta}\ \left\{p \frac{\partial}{\partial \theta}(\log p)\right\}\bigg|_{\theta=\theta^*}\ \mathrm{d}\boldsymbol{x} && (6)
\end{aligned}$$

また，$\mathbb{E}\big[\hat{\theta}\big] = \theta^*$ の両辺を θ^* で偏微分することを考える。微分と積分の順序が交換できるとき，

$$\begin{aligned}
1 &= \frac{\partial}{\partial \theta^*}\mathbb{E}\big[\hat{\theta}\big] && (7) \\
&= \frac{\partial}{\partial \theta^*} \int p\hat{\theta}\ \mathrm{d}\boldsymbol{x} && (8) \\
&= \int \frac{\partial}{\partial \theta^*}\big(p\hat{\theta}\big)\ \mathrm{d}\boldsymbol{x} && (9) \\
&= \int \hat{\theta}\frac{\partial p}{\partial \theta^*}\ \mathrm{d}\boldsymbol{x} && (10)
\end{aligned}$$

となる。いま，

$$p(\boldsymbol{x};\ \theta)\left(\frac{\partial}{\partial \theta} \log p(\boldsymbol{x};\ \theta)\right) = \frac{\partial}{\partial \theta} p(\boldsymbol{x}; \theta) \tag{11}$$

であるから，結局，

$$\int \hat{\theta}\ \left\{p \frac{\partial}{\partial \theta}(\log p)\right\}\bigg|_{\theta=\theta^*}\ \mathrm{d}\boldsymbol{x} = 1 \tag{12}$$

となる。

式 (20)，(31) から，

$$\mathbb{E}_{\boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_n \sim q(\cdot;\ \theta)} \left[\hat{\theta} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log q(\boldsymbol{x}_i;\ \theta) \bigg|_{\theta=\theta^*} \right] = \int \hat{\theta} \left\{ p \frac{\partial}{\partial \theta} (\log p) \right\} \bigg|_{\theta=\theta^*} \mathrm{d}\boldsymbol{x} = 1 \tag{13}$$

であることが示された。

宿題 2

微分と積分の順序が交換できるとき，フィッシャー情報行列について式 (15) が成立することを示す。

$$\boldsymbol{F}(\boldsymbol{\theta}) := \int \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q(\boldsymbol{x};\, \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q(\boldsymbol{x};\, \boldsymbol{\theta})\right)^{\mathrm{T}} q(\boldsymbol{x};\, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} \tag{14}$$

$$= -\int \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \log q(\boldsymbol{x};\, \boldsymbol{\theta})\right) q(\boldsymbol{x};\, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} \tag{15}$$

以下では，簡単のため $q(\boldsymbol{x};\, \boldsymbol{\theta})$ を q と記すこととする。$q, r \in \mathbb{R}$ について，

$$\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q\right) q = \frac{\partial}{\partial \boldsymbol{\theta}} q \tag{16}$$

$$\left(\frac{\partial r}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} = \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \tag{17}$$

が成立することから，

$$\boldsymbol{F}(\boldsymbol{\theta}) := \int \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q\right)^{\mathrm{T}} q\, \mathrm{d}\boldsymbol{x} \tag{18}$$

$$= \int \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q\right) q \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log q\right)^{\mathrm{T}} \mathrm{d}\boldsymbol{x} \tag{19}$$

$$= \int \frac{\partial q}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \log q\right) \mathrm{d}\boldsymbol{x} \tag{20}$$

となる。

いま，

$$G(\boldsymbol{\theta}) := \int \frac{\partial}{\partial \boldsymbol{\theta}} \left(q \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) \mathrm{d}\boldsymbol{x} \tag{21}$$

なる G について，微分と積分の順序が交換できるとき，

$$G(\boldsymbol{\theta}) = \int \frac{\partial}{\partial \boldsymbol{\theta}} \left(q \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) \mathrm{d}\boldsymbol{x} \tag{22}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\int q \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\, \mathrm{d}\boldsymbol{x}\right) \tag{23}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \left\{\int q \mathrm{d}\boldsymbol{x} \cdot \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} - \int \left(\int q \mathrm{d}\boldsymbol{x}\right) \cdot \frac{\partial}{\partial \boldsymbol{x}} \left(\frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) \mathrm{d}\boldsymbol{x}\right\} \tag{24}$$

が成立する。ここで，

$$\int q \mathrm{d}\boldsymbol{x} = 1 \tag{25}$$

から，

$$G(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} - \int \frac{\partial}{\partial \boldsymbol{x}} \left(\frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) \mathrm{d}\boldsymbol{x}\right) \tag{26}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} - \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) \tag{27}$$

$$= \boldsymbol{0} \tag{28}$$

となる。また，G について，次式も成立する。

$$G(\boldsymbol{\theta}) = \int \frac{\partial}{\partial \boldsymbol{\theta}} \left(q \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right) \mathrm{d}\boldsymbol{x} \tag{29}$$

$$= \int \left(\frac{\partial q}{\partial \boldsymbol{\theta}} \frac{\partial r}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + q \frac{\partial^2 r}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \right) \mathrm{d}\boldsymbol{x} \tag{30}$$

これらのことから，$r = \log q$ とすれば，

$$\int \frac{\partial q}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \log q \right) \mathrm{d}\boldsymbol{x} = -\int q \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \log q \right) \mathrm{d}\boldsymbol{x} \tag{31}$$

が成立することがわかる。

式 (20)，(31) から，

$$\boldsymbol{F}(\boldsymbol{\theta}) = -\int \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \log q \right) q \, \mathrm{d}\boldsymbol{x} \tag{32}$$

が示される。

宿題 3

二項分布から標本を発生させ，最尤推定量を求める。また，それを繰り返し，最尤推定量の分布を求める。標本数を変え，最尤推定量の漸近正規性の確認も行う。

理論

いま，二項分布では，

$$q(x;\ \theta) = {}_n\mathrm{C}_x \theta^x (1-\theta)^{n-x} \qquad (0 \le \theta \le 1) \tag{33}$$

であるから，その対数尤度および対数尤度の θ による微分は次のようになる。

$$\begin{aligned}
\log q(x;\ \theta) &= \log\left\{\frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{n-x}\right\} & (34)\\
&= \log(n!) - \log(x!) - \log((n-x)!) + x\log\theta + (n-x)\log(1-\theta) & (35)\\
\frac{\partial}{\partial\theta}(q(x;\ \theta)) &= \frac{x}{\theta} - \frac{n-x}{1-\theta} & (36)\\
&= \frac{x-n\theta}{\theta(1-\theta)} & (37)\\
&= \frac{1}{n}\frac{(x/n)-\theta}{\theta(1-\theta)} & (38)
\end{aligned}$$

これより，対数尤度を最大にするような θ，つまり θ の最尤推定量は，

$$\hat{\theta} = \frac{x}{n} \tag{39}$$

となる。したがって，

$$\frac{\partial}{\partial\theta}(q(x;\ \theta)) = \frac{1}{n}\frac{\hat{\theta}-\theta}{\theta(1-\theta)} \tag{40}$$

となる。また，これを用いると，フィッシャーの情報行列の期待値は次のようになる。

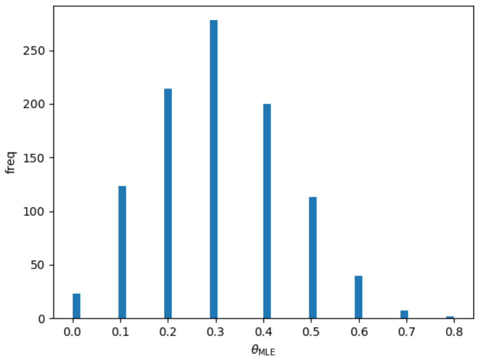
$$\begin{aligned}
\hat{F}(\theta^*) &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}(q(x;\ \theta))\bigg|_{\theta=\theta^*}\right)^2 & (41)\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\frac{\hat{\theta}-\theta^*}{\theta^*(1-\theta*)}\right)^2 & (42)
\end{aligned}$$

問題設定

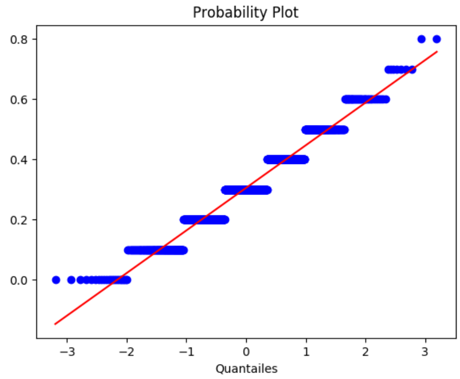
$\theta^* = 0.3$ であるような二項分布を用い，標本数 $n = 1\times 10^1$–1×10^5 について，それぞれ 1000 回最尤推定量を求めてその分布を求める。また，正規分布であるかどうかの判別にはシャピロウィルク検定を用いた。なお，プログラムは 9 ページの Listing 1 に示した。

表 1: 各サンプル数における最尤推定量の統計量

#Sample	mean	cov	F	$1/(nF)$	p	Normal Distribution
1×10^1	0.3055	2.086×10^{-2}	4.732	2.113×10^{-2}	3.371×10^{-17}	False
1×10^2	0.3011	2.200×10^{-3}	4.988	2.005×10^{-3}	3.770×10^{-4}	False
1×10^3	0.2998	2.083×10^{-4}	4.719	2.119×10^{-4}	0.1715	True
1×10^4	0.3001	1.891×10^{-5}	4.285	2.334×10^{-5}	0.6772	True
1×10^5	0.3001	2.175×10^{-6}	4.948	2.021×10^{-6}	0.4487	True



(a) ヒストグラム



(a) QQ プロット

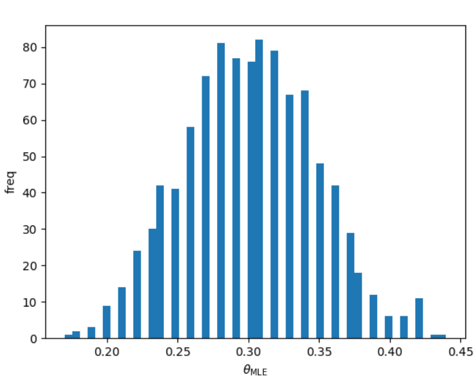
図 2: 標本数 10 のときの二項分布の最尤推定量のヒストグラムと QQ プロット

結果

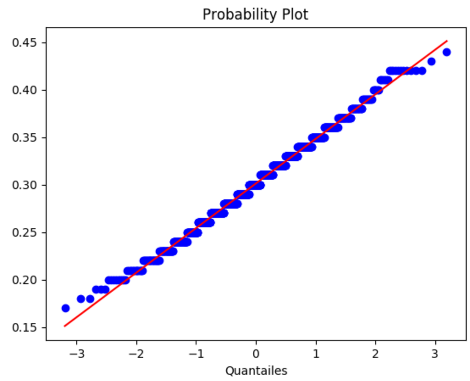
表 1 に，各標本数での結果を示す。なお，p はシャピロウィルク検定の p 値である。また，各標本数におけるヒストグラムと QQ プロットを図 2–10 に示す。

考察

標本数 n が 100 以下のときは，二項分布の最尤推定量の分布は正規分布とはみなされなかった。しかし，標本数を 1000 以上にしたところ，これは正規分布となった。これらのことから，最尤推定量の漸近正規性が確認された。

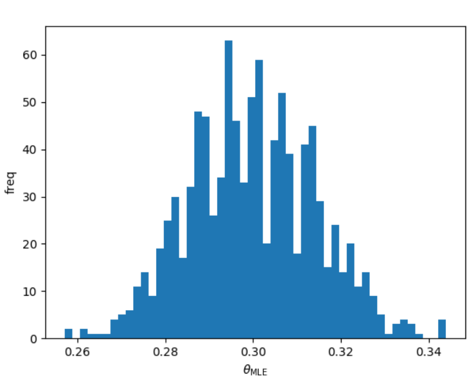


(a) ヒストグラム

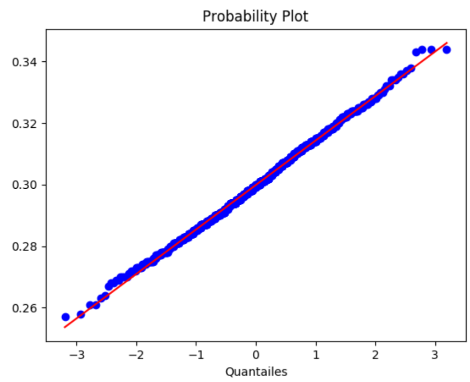


(a) QQ プロット

図 4: 標本数 100 のときの二項分布の最尤推定量のヒストグラムと QQ プロット

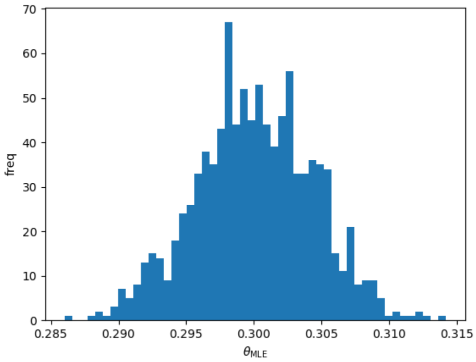


(a) ヒストグラム

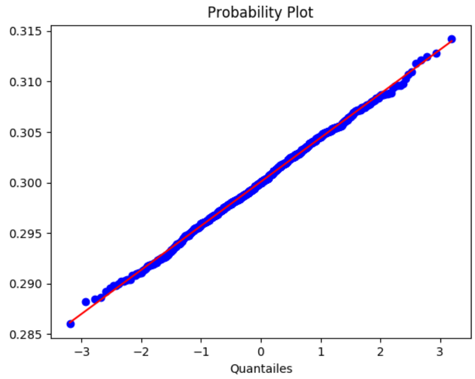


(a) QQ プロット

図 6: 標本数 1000 のときの二項分布の最尤推定量のヒストグラムと QQ プロット

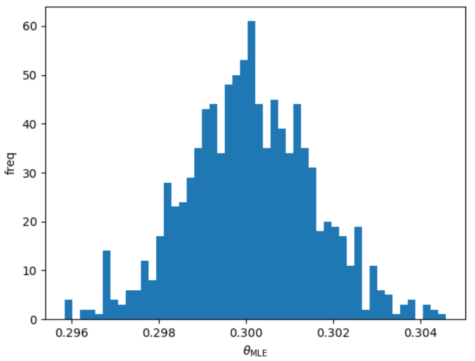


(a) ヒストグラム

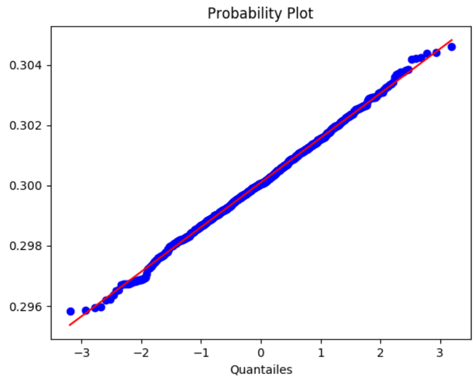


(a) QQ プロット

図 8: 標本数 10000 のときの二項分布の最尤推定量のヒストグラムと QQ プロット



(a) ヒストグラム



(a) QQ プロット

図 10: 標本数 100000 のときの二項分布の最尤推定量のヒストグラムと QQ プロット

プログラム

実行環境と用いた言語・ライブラリを以下の表 2 に示す。

表 2: プログラムの実行環境

OS	:	Microsoft Windows 10 Pro (64bit)
CPU	:	Intel(R) Core(TM) i5-4300U
RAM	:	4.00 GB
使用言語	:	Python3.6
可視化	:	matplotlib ライブラリ

Listings 1: assignment3.py

```python
# -*- coding: utf-8 -*-


import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats


def generate_sample(n, theta):
    sample = (np.random.rand(n) < theta)
    return sample


def main():
    # settings
    theta = 0.3
    n_sample = 100000
    n_mle = 1000
    np.random.seed(0)


    theta_mle_list = []
    fisher_matrices = []
    for _ in range(n_mle):
        sample = generate_sample(n_sample, theta)
        n_o = sample.sum()
        theta_mle = n_o / n_sample
        fisher_matrix = (1/n_sample) * ((n_sample/(theta*(1-theta))) * (theta_mle - theta))**2
        theta_mle_list.append(theta_mle)
        fisher_matrices.append(fisher_matrix)

```

```
    mean = np.mean(theta_mle_list)
    cov = np.cov(theta_mle_list)
    F = np.mean(fisher_matrices)

    W, p = stats.shapiro(theta_mle_list)
    is_normal_dist = (p > 0.05)

    print('n: {} \t n_MLE: {}'.format(n_sample, n_mle))
    print('True: theta*: {} \t 1/nF: {} \t F: {}'.format(theta, 1/(n_sample*F), F))
    print('MLE: mean: {:.4f} \t cov: {}'.format(mean, cov))
    print('is normal: {} \t W: {} \t p: {}'.format(is_normal_dist, W, p))


    # histgram
    plt.hist(theta_mle_list, bins=50)
    plt.xlabel(r'$\theta_\mathrm{MLE}$')
    plt.ylabel('freq')
    plt.savefig('../figures/hist_n{}.png'.format(n_sample))
    plt.show()

    # QQ plot
    stats.probplot(theta_mle_list, dist='norm', plot=plt)
    plt.xlabel('Quantailes')
    plt.ylabel('')
    plt.savefig('../figures/qqplot_n{}.png'.format(n_sample))
    plt.show()


if __name__ == '__main__':
    main()
```