宿題 1

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，

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

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。

$$\begin{aligned}
\log q(x;\ \theta) &= \log \left\{ \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} \right\} &(2)\\
&= \log(n!) - \log(x!) - \log((n-x)!) + x \log\theta + (n-x)\log(1-\theta) &(3)\\
\frac{\partial}{\partial\theta}(q(x;\ \theta)) &= \frac{x}{\theta} - \frac{n-x}{1-\theta} &(4)\\
&= \frac{x-n\theta}{\theta(1-\theta)} &(5)\\
&= \frac{1}{n}\frac{(x/n)-\theta}{\theta(1-\theta)} &(6)
\end{aligned}$$

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，

$$\hat{\theta} = \frac{x}{n} \tag{7}$$

となる。したがって，

$$\frac{\partial}{\partial\theta}(q(x;\ \theta)) = \frac{1}{n}\frac{\hat{\theta}-\theta}{\theta(1-\theta)} \tag{8}$$

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。

$$\begin{aligned}
\hat{F}(\theta^*) &= \frac{1}{N}\sum_{i=1}^{N}\left(\left.\frac{\partial}{\partial\theta}(q(x;\ \theta))\right|_{\theta=\theta^*} \right)^2 &(9)\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\frac{\hat{\theta}-\theta^*}{\theta^*(1-\theta*)} \right)^2 &(10)
\end{aligned}$$

問題設定

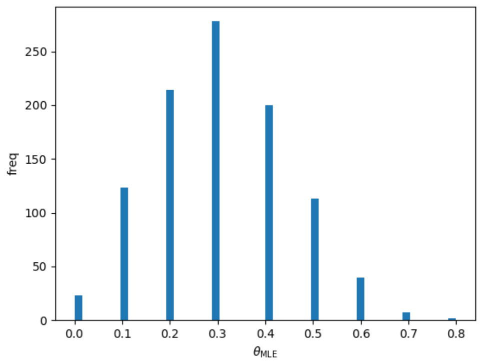
$\theta^* = 0.3$ であるような二項分布を用い，標本数 $n = 1\times 10^1$–1×10^5 について，それぞれ 1000 回最尤推定量を求めてその分布を求める。また，正規分布であるかどうかの判別にはシャピロウィルク検定を用いた。なお，プログラムは**??**ページの Listing **??**に示した。

結果

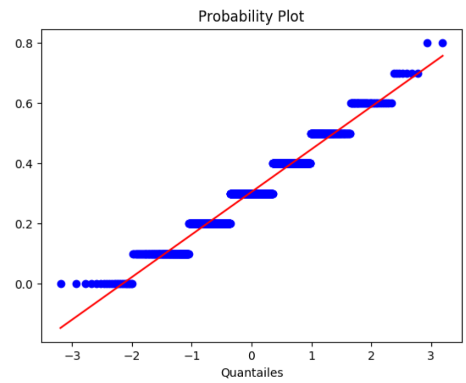
表 1 に，各標本数での結果を示す。なお，p はシャピロウィルク検定の p 値である。また，各標本数におけるヒストグラムと QQ プロットを図 1–5 に示す。

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

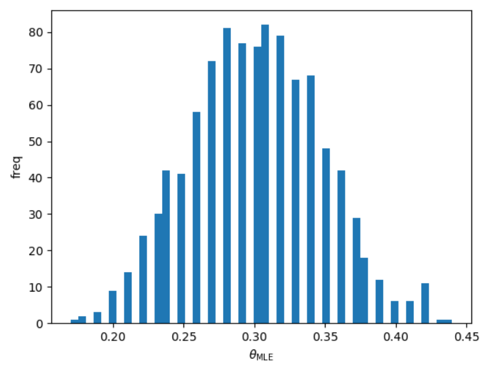


(a) ヒストグラム

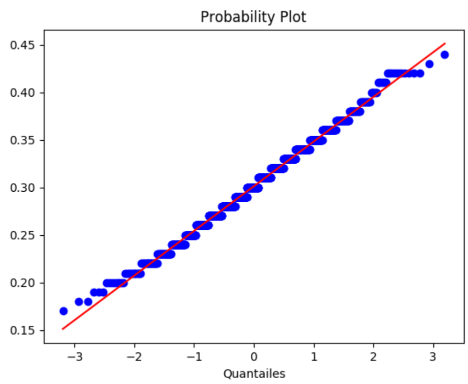


(b) QQ プロット

図 1: 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

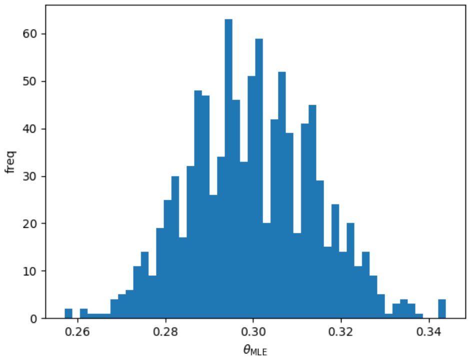


(a) ヒストグラム

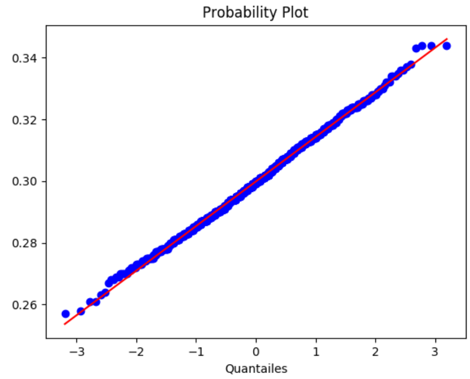


(b) QQ プロット

図 2: 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

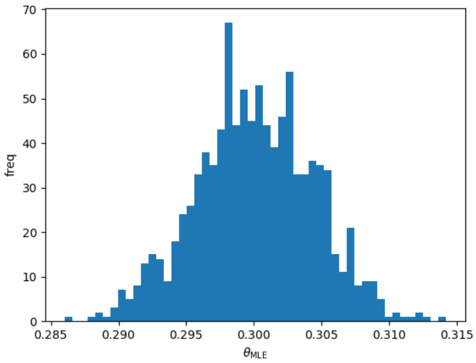


(a) ヒストグラム

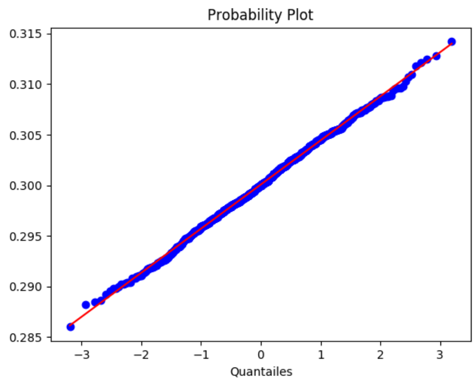


(b) QQ プロット

図 3: 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

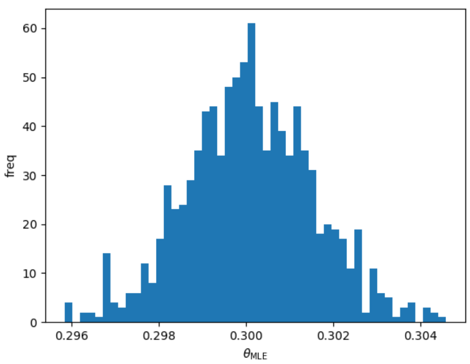


(a) ヒストグラム

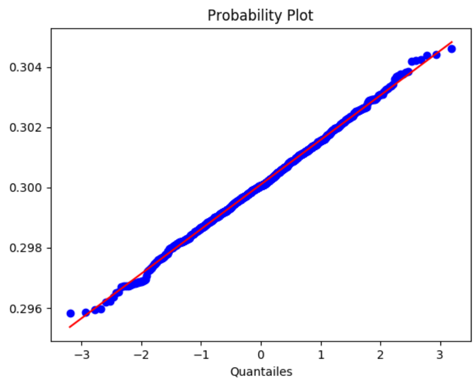


(b) QQ プロット

図 4: 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



(b) QQ プロット

図 5: 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

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。