経済学特論（経済時系列分析入門）

Tue., 8:50-10:20

Room # 4 (法経講義棟)

- The prerequisite of this class is **Basic Statistics (統計基礎)** (by Prof. Fukushige, Tue., 16:20-17:50, this semester), **Special Lectures in Economics (Statistical Analysis), 経済学特論（統計解析）** (by Prof. Fukushige, Tue., 14:40-16:10, this semester), **Econometrics (エコノメトリックス)** (undergraduate level, next semester, 『計量経済学』山本 拓 著，新世社), **Econometrics I (計量経済 I)** (graduate level, this semester), and **Econometrics II (計量経済 II)** (graduate level, next semester).

代表的テキスト：

・J.D. Hamilton (1994) *Time Series Analysis*

沖本・井上訳 (2006)『時系列解析 (上・下)』

・A.C. Harvey (1981) *Time Series Models*

国友・山本訳 (1985)『時系列モデル入門』

・沖本竜義 (2010)『経済・ファイナンスデータの計量時系列分析』

1　最尤法 — 復習

n 個の確率変数 $X_1, X_2, \cdots, X_n$ は互いに独立で，同じ確率分布 $f(x) \equiv f(x;\theta)$ とする。ただし，θ は母数で，例えば，$\theta = (\mu, \sigma^2)$ である。

$X_1, X_2, \cdots, X_n$ の結合分布は，互いに独立なので，

$$f(x_1, x_2, \cdots, x_n; \theta) \equiv \prod_{i=1}^{n} f(x_i; \theta)$$

と表される。

観測データ $x_1, x_2, \cdots, x_n$ を与えたもとで，$\prod_{i=1}^{n} f(x_i; \theta)$ は θ の関数として表される。すなわち，

$$l(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

となる。

$l(\theta)$ を尤度関数と呼ぶ。

$$\max_{\theta} \ l(\theta)$$

となる θ を最尤推定値 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ と呼ぶ。

データ $x_1, x_2, \cdots, x_n$ を確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ を最尤推定量と呼ぶ。

$$\max_{\theta} \ l(\theta)$$

と

$$\max_{\theta} \ \log l(\theta)$$

の θ の解はともに同じものであることに注意。$\log l(\theta)$ を対数尤度関数と呼ぶ。

最尤推定量の性質：　n が大きいとき,

$$\hat{\theta} \sim N(\theta, \sigma_\theta^2)$$

ただし,

$$\sigma_\theta^2 = \frac{1}{\sum_{i=1}^n \mathrm{E}\Big[\Big(\dfrac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\Big)^2\Big]} = -\frac{1}{\sum_{i=1}^n \mathrm{E}\Big[\dfrac{\mathrm{d}^2\log f(X_i;\theta)}{\mathrm{d}\theta^2}\Big]}$$

θ がベクトル $(k \times 1)$ の場合，n が大きいとき,

$$\hat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし,

$$\Sigma_\theta = \Big(\sum_{i=1}^n \mathrm{E}\Big[\Big(\frac{\partial\log f(X_i;\theta)}{\partial\theta}\Big)\Big(\frac{\partial\log f(X_i;\theta)}{\partial\theta}\Big)'\Big]\Big)^{-1} = -\Big(\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial^2\log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}$$

例 1：　正規母集団 N(μ, σ^2) からの標本値 $x_1, x_2, \cdots, x_n$ を用いて,

(1) σ^2 が既知のとき，μ の最尤推定値と最尤推定量

(2) σ^2 が未知のとき，μ と σ^2 の最尤推定値と最尤推定量

をそれぞれ求める。

［解］ N(μ, σ^2) の密度関数は,

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

となる。したがって，互いに独立な $X_1, X_2, \cdots, X_n$ の結合分布は,

$$f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) \equiv \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$

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

となる。

(1) σ^2 が既知のとき，尤度関数 $l(\mu)$ は,

$$l(\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\Bigr)$$

となる。

$l(\mu)$ を最大にする μ と $\log l(\mu)$ を最大にする μ は同じになる。

したがって，対数尤度関数は,

$$\log l(\mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

となる。この対数尤度関数を μ に関して最大化すると,

$$\frac{\mathrm{d}\log l(\mu)}{\mathrm{d}\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

となる μ を求める。μ の解を $\hat{\mu}$ とすると，μ の最尤推定値は,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i \equiv \overline{x}$$

を得る。

さらに，観測値 $x_1, x_2, \cdots, x_n$ をその確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，μ の最尤推定量は，

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \overline{X}$$

となる。

$\hat{\mu}$ の分散を求めるために，密度関数の対数を取って，

$$\log f(X_i;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X_i - \mu)^2$$

となり，μ に関して微分して，

$$\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu} = \frac{1}{\sigma^2}(X_i - \mu)$$

が得られる。両辺を二乗して，

$$\left(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\right)^2 = \frac{1}{\sigma^4}(X_i - \mu)^2$$

となり，期待値を取ると，

$$\mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\Big)^2\Big] = \frac{1}{\sigma^4}\mathrm{E}[(X_i-\mu)^2] = \frac{1}{\sigma^2}$$

と計算される。

最尤推定量の性質から，n が大きいとき，

$$\hat{\mu} \sim N(\mu, \sigma_\mu^2)$$

ただし，

$$\sigma_\mu^2 = \frac{1}{\sum_{i=1}^n \mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\Big)^2\Big]} = \frac{\sigma^2}{n}$$

この場合は，n の大きさに関わらず，$\hat{\mu} \sim N(\mu, \sigma_\mu^2)$ が成り立つ。

(2) σ^2 が未知のとき，μ と σ^2 の尤度関数は，

$$l(\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\Big)$$

となる。

対数尤度関数は,

$$\log l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

と表される。

μ と σ^2 について，最大化するためには,

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

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

の連立方程式を解く。

μ, σ^2 の解を $\hat{\mu}, \hat{\sigma}^2$ とすると，最尤推定値は,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}) \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})$$

となる。

観測値 $x_1, x_2, \cdots, x_n$ をその確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，μ, σ^2 の最尤推定量は,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}) \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})$$

となる。

σ^2 の最尤推定量 $\hat{\sigma}^2$ は，σ^2 の不偏推定量 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ とは異なることに注意。

$\theta = (\mu, \sigma^2)'$ とする。n が大きいとき,

$$\hat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし,

$$\Sigma_\theta = -\Big(\sum_{i=1}^{n} \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}$$

となる。Σ_θ を得るために，密度関数の対数を取って,

$$\log f(X_i;\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(X_i-\mu)^2$$

とばり，μ，σ^2 について微分すると,

$$\frac{\partial \log f(X_i;\theta)}{\partial\theta} = \begin{pmatrix} \dfrac{\partial \log f(X_i;\theta)}{\partial\mu} \\ \dfrac{\partial \log f(X_i;\theta)}{\partial\sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma^2}(X_i-\mu) \\ -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(X_i-\mu)^2 \end{pmatrix}$$

となる。さらに，もう一度，微分すると，

$$\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 \log f(X_i;\theta)}{\partial\mu^2} & \frac{\partial^2 \log f(X_i;\theta)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 \log f(X_i;\theta)}{\partial\sigma^2\partial\mu} & \frac{\partial^2 \log f(X_i;\theta)}{\partial(\sigma^2)^2} \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{1}{\sigma^4}(X_i-\mu) \\ -\frac{1}{\sigma^4}(X_i-\mu) & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X_i-\mu)^2 \end{pmatrix}$$

となる。期待値を取って，

$$\mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big] = \begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{1}{\sigma^4}\mathrm{E}(X_i-\mu) \\ -\frac{1}{\sigma^4}\mathrm{E}(X_i-\mu) & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\mathrm{E}[(X_i-\mu)^2] \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{1}{2\sigma^4} \end{pmatrix}$$

を得る。

よって，

$$\Sigma_\theta = -\Big(\sum_{i=1}^{n} \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1} = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

が得られる。

まとめると，μ，σ^2 の最尤推定量 $\hat{\mu} = (1/n)\sum_{i=1}^{n} X_i$，$\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}(X_i - \overline{X})^2$ の分布は，n が大きいとき，

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}\right)$$

となる。

例 2： $X_1, X_n, \cdots, X_n$ は互いに独立で，それぞれパラメータ p を持ったベルヌイ分布に従うものとする。すなわち，X_i の確率関数は,

$$f(x;p) = p^x(1-p)^{1-x} \qquad\qquad x = 0, 1$$

となる。

このとき尤度関数は,

$$l(p) = \prod_{i=1}^{n} f(x_i;p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

となり，対数尤度関数は,

$$\begin{aligned}\log l(p) &= \sum_{i=1}^{n} \log f(x_i;p) \\ &= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i)\end{aligned}$$

$$= \log(p) \sum_{i=1}^{n} x_i + \log(1-p)(n - \sum_{i=1}^{n} x_i)$$

となる。

$\log l(p)$ を最大にする p を求める。

$$\frac{\mathrm{d} \log l(p)}{\mathrm{d} p} = \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} x_i) = 0$$

したがって，p について解くと，p の最尤推定値 $\hat{p}$ は,

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

となる。

さらに，x_i を X_i で置き換えて，p の最尤推定量 $\hat{p}$ は,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

となる。

$\hat{p}$ の分布を求める。確率関数の対数を取って，

$$\log f(X_i; p) = X_i \log(p) + (1 - X_i) \log(1 - p)$$

となる。p について微分すると，

$$\frac{\mathrm{d} \log f(X_i; p)}{\mathrm{d}p} = \frac{X_i}{p} - \frac{1 - X_i}{1 - p} = \frac{X_i - p}{p(1 - p)}$$

となる。さらに，両辺を二乗して，期待値を取ると，

$$\mathrm{E}\Big[\Big(\frac{\mathrm{d} \log f(X_i; p)}{\mathrm{d}p}\Big)^2\Big] = \frac{\mathrm{E}[(X_i - p)^2]}{p^2(1 - p)^2}$$

となる。期待値を計算すると，

$$\begin{aligned}\mathrm{E}[(X_i - p)^2] &= \sum_{x_i=0}^{1} (x_i - p)^2 f(x_i; p) = \sum_{x_i=0}^{1} (x_i - p)^2 p^{x_i} (1 - p)^{1-x_i} \\ &= p^2(1 - p) + (1 - p)^2 p = p(1 - p)\end{aligned}$$

となるので,

$$\sigma_\theta^2 = \frac{1}{\sum_{i=1}^n \mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i; p)}{\mathrm{d}p}\Big)^2\Big]} = \frac{p(1-p)}{n}$$

が得られる。したがって,$\hat{p}$ の分布は,n が大きいとき,

$$\hat{p} \sim N(p,\ \frac{p(1-p)}{n})$$

となる。

例 3： $X_1, X_n, \cdots, X_n$ は互いに独立で，それぞれパラメータ λ を持ったポアソン分布に従うものとする。すなわち，X_i の確率関数は,

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad\qquad x = 0, 1, 2, \cdots$$

となる。

このとき尤度関数は,

$$l(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

となり，対数尤度関数は,

$$\log l(\lambda) = \sum_{i=1}^{n} \log f(x_i;\lambda) = \log(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log(x_i!)$$

となる。

$\log l(\lambda)$ を最大にする p を求める。

$$\frac{\mathrm{d}\log l(\lambda)}{\mathrm{d}\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

したがって，λ について解くと，λ の最尤推定値 $\hat{\lambda}$ は,

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

となる。

さらに，x_i を X_i で置き換えて，λ の最尤推定量 $\hat{\lambda}$ は,

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

となる。

$\hat{\lambda}$ の分布を求める。確率関数に対数を取って,

$$\log f(X_i;\lambda) = X_i \log(\lambda) - \lambda - \log(X_i!)$$

となる。λ に関して，微分すると，

$$\frac{\mathrm{d}\log f(X_i;\lambda)}{\mathrm{d}\lambda} = \frac{X_i}{\lambda} - 1$$

を得る。再度，λ に関して，微分すると，

$$\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2} = -\frac{X_i}{\lambda^2}$$

となる。期待値をとって，

$$\mathrm{E}\Big(\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Big) = \frac{\mathrm{E}(X_i)}{\lambda^2}$$

が得られる。期待値を計算すると，

$$\begin{aligned}\mathrm{E}(X_i) &= \sum_{x=0}^{\infty} x f(x;\lambda) = \sum_{x=0}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\ &= \sum_{x=1}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \lambda\frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \sum_{x=0}^{\infty} \lambda\frac{\lambda^x e^{-\lambda}}{x!} = \lambda\end{aligned}$$

となる。したがって,

$$\sigma_\theta^2 = -\frac{1}{\sum_{i=1}^n \mathrm{E}\Bigl(\frac{\mathrm{d}^2 \log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Bigr)} = \frac{\lambda}{n}$$

となる。よって,$\hat{\lambda}$ の分布は,n が大きいとき,

$$\hat{\lambda} \sim N(\lambda,\ \frac{\lambda}{n})$$

を得る。

例 4： $X_1, X_n, \cdots, X_n$ は互いに独立で，それぞれパラメータ λ を持った指数分布に従うものとする。すなわち，X_i の密度関数は,

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad\qquad x > 0$$

となる。

このとき尤度関数は,

$$l(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

となり，対数尤度関数は,

$$\log l(\lambda) = \sum_{i=1}^{n} \log f(x_i;\lambda) = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

となる。

$\log l(\lambda)$ を最大にする p を求める。

$$\frac{\mathrm{d}\log l(\lambda)}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

したがって，λ について解くと，λ の最尤推定値 $\hat{\lambda}$ は,

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

となる。

さらに，x_i を X_i で置き換えて，λ の最尤推定量 $\hat{\lambda}$ は,

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

となる。

$\hat{\lambda}$ の分布を求める。密度関数の対数を取って,

$$\log f(X_i; \lambda) = \log \lambda - \lambda X_i$$

となる。λ について微分すると，

$$\frac{\mathrm{d}\log f(X_i;\lambda)}{\mathrm{d}\lambda} = \frac{1}{\lambda} - X_i$$

を得る。再度，微分して，

$$\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2} = -\frac{1}{\lambda^2}$$

が得られる。

$$\sigma_\theta^2 = -\frac{1}{\sum_{i=1}^n \mathrm{E}\Big(\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Big)} = \frac{\lambda^2}{n}$$

となるので，n が大きいとき，

$$\hat{\lambda} \sim N(\lambda,\ \frac{\lambda^2}{n})$$

を得る。