平成 25 年度統計解析手法期末試験 解答例

中田雄大

2014 年 7 月 12 日

問 1　パラメータ推定

(1) 母集団が従う分布のパラメータを θ とすると，同時確率密度関数は $\boldsymbol{x} = (x_1, x_2, x_3, \ldots, x_n)$ を確率変数の観測結果として $f_\theta(\boldsymbol{x})$ と書ける．いま，観測の結果によって $\boldsymbol{x}$ が定まったとき，θ の値は様々に考えられるので，$f_\theta(\boldsymbol{x})$ を逆に θ を変数とする尤度関数 $L_{\boldsymbol{x}}(\theta)$ と考えると，これは θ の値によって観測結果 $\boldsymbol{x}$ を得る確率の変化を示す関数となる．同じ観測結果に対して $L_{\boldsymbol{x}}(\theta_1) > L_{\boldsymbol{x}}(\theta_2)$ であれば母集団のパラメータが θ_1 である分布に従っていると考えるのが妥当である．このようにして，$L_{\boldsymbol{x}}(\theta)$ の最大値を与える $\theta = \hat{\theta}$ をパラメータとする母集団分布に従っていると考える方法が尤度法である．

(2) 対数尤度関数 $\tilde{L}(\mu, \sigma^2) = \ln L(\mu, \sigma^2)$ の最大値を考える．問題で与えられた尤度関数より

$$\begin{cases} \dfrac{\partial \tilde{L}}{\partial \mu} = \dfrac{1}{\sigma^2} \sum (x_i - \mu) & (1.1) \\ \dfrac{\partial \tilde{L}}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^2} \sum (x_i - \mu)^2. & (1.2) \end{cases}$$

$(\mu, \sigma^2) = (\hat{\mu}, \hat{\sigma}^2)$ のときに最大値をとるとすると，$\bar{x} = \dfrac{1}{n} \sum x_i$ として式 (1.1), 式 (1.2) より

$$\begin{cases} \dfrac{1}{\hat{\sigma}^2} \sum (z_i - \hat{\mu}) = 0 \\ -\dfrac{n}{2\hat{\sigma}^2} + \dfrac{1}{2\hat{\sigma}^2} \sum (x_i - \hat{\mu})^2 = 0 \end{cases}$$

$$\therefore\ \hat{\theta} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2. \tag{1.3}$$

問 2　区間推定

(1) 標本平均，標本数をそれぞれ $\bar{X}, n$ として $\bar{X} \sim N(\mu, \sigma^2/n)$.

(2) 母分散 σ^2 が既知であるから標準化すると $z = \dfrac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$ となる．ここに，$n = 100$ である．これを $P(-z_{0.005} \le z \le z_{0.005}) = 0.99$ に代入して

$$P(-z_{0.005} \le \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} \le z_{0.005}) = 0.99. \tag{2.1}$$

よって，μ の 99% 信頼区間は $[18.985, 20.015]$ となる．

(3) 一般に母集団分布が $N(\mu, \sigma^2)$ であるような確率変数の標本平均 $\bar{X}$ に関して，標本数，有意水準をそれぞれ n, α とすると

$$P(-z_{\frac{\alpha}{2}} \leq \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}) = 1-\alpha \tag{2.2}$$

$$\therefore\ P\left(\bar{X} - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1-\alpha. \tag{2.3}$$

式 (2.3) より母平均 μ の $100(1-\alpha)\%$ 信頼区間は $[\bar{X} - \sigma z_{\frac{\alpha}{2}}/\sqrt{n}, \bar{X} + \sigma z_{\frac{\alpha}{2}}/\sqrt{n}]$ となる．よって，標本数が大きいほど，また標準偏差が小さいほど母平均の信頼区間の区間幅は小さくなり推定の精度は上昇する．

問 3　最小二乗法

(1) 最小二乗法においては誤差項の二乗和 $S = \sum[Y_i - (\beta_0 + \beta_1 X_i)]^2$ を最小化する．S が最小となる条件は S を各回帰係数で偏微分して

$$\frac{\partial S}{\partial \beta_0} = \frac{\partial S}{\partial \beta_1} = 0 \tag{3.1}$$

$$\therefore \begin{cases} \sum(Y_i - \beta_0 - \beta_1 X_i) = 0 & (3.2) \\ \sum X_i(Y_i - \beta_0 - \beta_1 X_i) = 0 & (3.3) \end{cases}$$

となる．式 (3.2) と式 (3.3) が正規方程式である．

(2) (1) より正規方程式を解いて最小二乗推定量は

$$(\hat{\beta}_0, \hat{\beta}_1) = \left(\bar{Y} - \hat{\beta}_1 \bar{X}, \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}\right). \tag{3.4}$$

ここに $\bar{X}, \bar{Y}$ はそれぞれ X, Y の標本平均である．

(3) 通常の最小二乗法が適用可能であるための仮定条件の一つに $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \sigma^2 I$ というものがある．ここに，$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}}$ であり，I は単位行列である．これは誤差項の分散は観測値によらないことと異なる観測値の誤差間に相関関係はないことを仮定しており，このとき最小二乗法で求めた推定量は最良線形不偏推定量であることが Gauß–Markov の定理により保証されている．逆に分散共分散行列が単位行列の正の定数倍でないときに通常の最小二乗法を用いると得られる推定値は一致推定量ではあるが最良不偏推定量ではない．このようなときに最良不偏推定量を求めるための方法が一般化最小二乗法である．

問 4　主成分分析

(1) 主成分分析では変数間に相関関係が存在する観測値について直交変換により主成分と呼ばれる相関関係のない変数に変換する．これによってより少ない数の変数（軸）でモデル化・記述することが出来る．

(2) 分散共分散行列を Σ とおく．Σ の固有値を求めると特性方程式は，

$$\det(\lambda I - \Sigma) = 0$$
$$\therefore \begin{vmatrix} \lambda - 4 & -\sqrt{2} \\ -\sqrt{2} & \lambda - 3 \end{vmatrix} = 0. \tag{4.1}$$

これを解いて固有値として $\lambda_1 = 5,\ \lambda_2 = 2$ を得る．これより第一主成分の寄与率は

$$\frac{5}{5+2} \times 100\% \simeq 71.4\%. \tag{4.2}$$

(3) 主成分分析は観測値に直交変換を施すことにより変数間に相関のない新たな変数の軸でデータを見るための手法であり，因子分析は観測値がどのような主成分（共通因子）によって定まるのかを見つける手法である．