となる。

対数尤度関数は,

$$\log l(\alpha,\beta,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2$$

となる。

$\log l(\alpha,\beta,\sigma^2)$ を最大にするために,

$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i) = 0$$

$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i(Y_i - \alpha - \beta X_i) = 0$$

$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2 = 0$$

の連立方程式を解く。

上 2 つの式は σ^2 に依存していない。α，β の最尤推定量は最小二乗推定量と同じになる。

すなわち，

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

σ^2 の最尤推定量は，

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

となり，s^2 とは異なる。

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

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

ただし,

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

Y_i の密度関数 $g(Y_i;\theta)$ の対数は,

$$\log g(Y_i;\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2$$

となる。一階微分は,

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

となる。二階微分は，

$$
\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha^2} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha\partial\beta} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha\partial\sigma^2} \\ \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta\partial\alpha} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta^2} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta\partial\sigma^2} \\ \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\sigma^2\partial\alpha} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\sigma^2\partial\beta} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial(\sigma^2)^2} \end{pmatrix}
$$

$$
= \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{u_i}{\sigma^4} \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_i u_i}{\sigma^4} \\ -\dfrac{u_i}{\sigma^4} & -\dfrac{X_i u_i}{\sigma^4} & \dfrac{1}{2\sigma^4} - \dfrac{u_i^2}{\sigma^6} \end{pmatrix}
$$

ただし，$u_i = Y_i - \alpha - \beta X_i$

期待値をとると，

$$
\mathrm{E}\Big(\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'}\Big) = \mathrm{E}\begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{u_i}{\sigma^4} \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_i u_i}{\sigma^4} \\ -\dfrac{u_i}{\sigma^4} & -\dfrac{X_i u_i}{\sigma^4} & \dfrac{1}{2\sigma^4} - \dfrac{u_i^2}{\sigma^6} \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & 0 \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & 0 \\ 0 & 0 & -\dfrac{1}{2\sigma^4} \end{pmatrix}
$$

となる。

したがって,

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

$$= \begin{pmatrix} \sigma^2 \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

したがって,

$$\begin{pmatrix} \hat\alpha \\ \hat\beta \end{pmatrix} \sim N\Big(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \sigma^2 \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1}\Big)$$

となる。

$\longrightarrow$ 最小二乗推定量の分布と同じ。

5　尤度比検定

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。

尤度関数は,

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

となる。

θ の制約つき最尤推定量を $\widetilde{\theta}$，制約無し最尤推定量を $\hat{\theta}$ とする。

制約の数を G 個とする。

$\dfrac{l(\widetilde{\theta})}{l(\hat{\theta})}$ を尤度比と呼ぶ

検定方法 1：　尤度比がある値より小さいときに，帰無仮説を棄却する。すなわち,

$$\frac{l(\widetilde{\theta})}{l(\hat{\theta})} < c$$

となるときに，帰無仮説を棄却する。この場合，c を次のようにして求める必要がある。

$$\int \cdots \int \prod_{i=1}^{n} f(x_i; \widetilde{\theta}) \mathrm{d}x_1 \cdots \mathrm{d}x_n = \alpha$$

ただし，α は有意水準（帰無仮説が正しいときに，帰無仮説を棄却する確率）を表す。

検定方法 2（大標本検定）：　または，$n \longrightarrow \infty$ のとき，

$$-2 \log \frac{l(\widetilde{\theta})}{l(\hat{\theta})} \longrightarrow \chi^2(G)$$

となる。

この検定を尤度比検定と呼ぶ。

例 1：　正規母集団 $\mathrm{N}(\mu, \sigma^2)$ からの標本値 $x_1, x_2, \cdots, x_n$ を用いて，σ^2 が既知のとき，帰無仮説 H_0：$\mu = \mu_0$，H_1：$\mu \neq \mu_0$ の尤度比検定を行う。

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。

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。

尤度比検定統計量は,

$$\frac{l(\mu_0)}{l(\overline{X})} = \frac{\exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2\Bigr)}{\exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2\Bigr)} = \exp\Bigl(-\frac{1}{2\sigma^2/n}(\overline{X} - \mu_0)^2\Bigr) < c$$

となる c を求める。

H_0 が正しいときに, $\sqrt{n}(\overline{X}-\mu_0)/\sigma \sim N(0,1)$ となるので,

$$P\Big(\Big|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\alpha/2}\Big) = \alpha$$

すなわち,

$$P\Big(\exp\Big(-\frac{1}{2\sigma^2/n}(\overline{X}-\mu_0)^2\Big) < \exp\Big(-\frac{1}{2}z_{\alpha/2}^2\Big)\Big) = \alpha$$

と変形できる。したがって,

$$c = \exp\Big(-\frac{1}{2}z_{\alpha/2}^2\Big)$$

とすればよい。

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,

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

となる。

このとき尤度関数は,

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

となる。

p の最尤推定量 $\hat{p}$ は,

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

である。

次の仮説検定を考える。

$$H_0: \ p = p_0 \qquad\qquad H_1: \ p \neq p_0$$

$\rightarrow$ 制約数は 1 つ。$(G = 1)$

尤度比は,

$$\frac{l(p_0)}{l(\hat{p})} = \frac{\prod_{i=1}^{n} p_0^{X_i}(1-p_0)^{1-X_i}}{\prod_{i=1}^{n} \hat{p}^{X_i}(1-\hat{p})^{1-X_i}}$$

したがって，$n \longrightarrow \infty$ のとき,

$$-2\log\frac{l(p_0)}{l(\hat{p})} = -2\log\frac{p_0}{\hat{p}}\sum_{i=1}^{n} X_i - 2\log\frac{1-p_0}{1-\hat{p}}\sum_{i=1}^{n}(1-X_i) \longrightarrow \chi^2(1)$$

$\chi^2(1)$ 分布の上側 100 α% 点を $\chi^2_\alpha(1)$ とするとき,

$$-2\log\frac{l(p_0)}{l(\hat{p})} > \chi^2_\alpha(1)$$

のとき，帰無仮説 H_0： $p = p_0$ を棄却する。

例 3： 回帰モデル

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

$$u_i \sim N(0, \sigma^2) \qquad i = 1, 2, \cdots, n$$

について，$\beta_1, \cdots, \beta_k$ に関する仮説の尤度比検定を行う。

例えば,

$$H_0:\ \beta_1 = 0$$

$$H_0:\ \beta_1 + \beta_2 = 1$$

$$H_0:\ \beta_1 = \beta_2 = \beta_3 = 0$$

などのような仮説検定

$\theta = (\beta_1, \cdots, \beta_k, \sigma^2)$ とする。

尤度関数は,

$$l(\theta) = \prod_{i=1}^{n} f(Y_i;\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 X_{1i} - \cdots - \beta_k X_{ki})^2\right)$$

となる。

H_0 の制約つき最尤推定量を $\widetilde{\theta} = (\widetilde{\beta}_1, \cdots, \widetilde{\beta}_k, \widetilde{\sigma}^2)$ とする。この仮設に含まれる制約数を G とする。

制約なし最尤推定量を $\hat{\theta} = (\hat{\beta}_1, \cdots, \hat{\beta}_k, \hat{\sigma}^2)$ とする。

尤度比

$$\frac{l(\widetilde{\theta})}{l(\hat{\theta})} = \frac{(2\pi\widetilde{\sigma}^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\widetilde{\sigma}^2}\sum_{i=1}^{n}(Y_i - \widetilde{\beta}_1 X_{1i} - \cdots - \widetilde{\beta}_k X_{ki})^2\Bigr)}{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki})^2\Bigr)}$$

$$= \frac{(\widetilde{\sigma}^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{n-G}{2}\Bigr)}{(\hat{\sigma}^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{n-k}{2}\Bigr)} = \left(\frac{\frac{1}{n-G}\sum_{i=1}^{n}\widetilde{u}_i^2}{\frac{1}{n-k}\sum_{i=1}^{n}\hat{u}_i^2}\right)^{-n/2} \exp\Bigl(-\frac{k-G}{2}\Bigr)$$

$$= \exp\Bigl(-\frac{k-G}{2}\Bigr)\left(\frac{n-k}{n-G}\right)^{-n/2}\left(\frac{\sum_{i=1}^{n}\widetilde{u}_i^2}{\sum_{i=1}^{n}\hat{u}_i^2}\right)^{-n/2}$$

$$
= \exp\Bigl(-\frac{k-G}{2}\Bigr)\left(\frac{n-k}{n-G}\right)^{-n/2}\left(1+\frac{\sum_{i=1}^n \widetilde{u}_i^2-\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n \hat{u}_i^2}\right)^{-n/2}
$$

$$
= \exp\Bigl(-\frac{k-G}{2}\Bigr)\left(\frac{n-k}{n-G}\right)^{-n/2}\left(1+\frac{G}{n-k}\frac{(\sum_{i=1}^n \widetilde{u}_i^2-\sum_{i=1}^n \hat{u}_i^2)/G}{\sum_{i=1}^n \hat{u}_i^2/(n-k)}\right)^{-n/2} < c
$$

のとき仮説を棄却する。

$$
\frac{(\sum_{i=1}^n \widetilde{u}_i^2-\sum_{i=1}^n \hat{u}_i^2)/G}{\sum_{i=1}^n \hat{u}_i^2/(n-k)} \sim F(G, n-k)
$$

を利用すると c が求まる。

ただし，途中で以下を利用

$$
\widetilde{\sigma}^2 = \frac{1}{n-G}\sum_{i=1}^n (Y_i-\widetilde{\beta}_1 X_{1i}-\cdots-\widetilde{\beta}_k X_{ki})^2 = \frac{1}{n-G}\sum_{i=1}^n \widetilde{u}_i^2
$$

$$
\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^n (Y_i-\hat{\beta}_1 X_{1i}-\cdots-\hat{\beta}_k X_{ki})^2 = \frac{1}{n-k}\sum_{i=1}^n \hat{u}_i^2
$$

近似的には,

$$
\begin{aligned}
-2\log\frac{l(\widetilde{\theta})}{l(\hat{\theta})} &= -2\log\frac{(\widetilde{\sigma}^2)^{-\frac{n}{2}}\exp\left(-\dfrac{n-G}{2}\right)}{(\hat{\sigma}^2)^{-\frac{n}{2}}\exp\left(-\dfrac{n-k}{2}\right)} \\
&= n\log\left(\frac{\widetilde{\sigma}^2}{\hat{\sigma}^2}\right) + (k-G) \ \longrightarrow\ \chi^2(G)
\end{aligned}
$$

6 Time Series Analysis (時系列分析)

6.1 Introduction

1. **Stationarity (定常性) :**

   Let $y_1, y_2, \cdots, y_T$ be time series data.

   (a) **Weak Stationarity (弱定常性) :**

$$\mathrm{E}(y_t) = \mu,$$

$$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

   The first moment does not depend on time.

   The second moment depends only on time difference.

(b) **Strong Stationarity (強定常性) :**

Let $f(y_{t_1}, y_{t_2}, \cdots, y_{t_r})$ be the joint distribution of $y_{t_1}, y_{t_2}, \cdots, y_{t_r}$.

$$f(y_{t_1}, y_{t_2}, \cdots, y_{t_r}) = f(y_{t_1+\tau}, y_{t_2+\tau}, \cdots, y_{t_r+\tau})$$

All the moments are same for all τ.

2. **Ergodicity (エルゴード性) :**

As time difference between two data is large, the two data become independent.

$y_1, y_2, \cdots, y_T$ is said to be ergodic in mean when $\bar{y}$ converges in probability to $\mathrm{E}(y_t)$.

3. **Auto-covariance Function (自己共分散関数) :**

$$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

$$\gamma(\tau) = \gamma(-\tau)$$

4. **Auto-correlation Function (自己相関関数) :**

$$\rho(\tau) = \frac{\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu))}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Note that $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-\tau}) = \gamma(0)$.

5. **Sample Mean (標本平均) :**

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

6. **Sample Auto-covariance (標本自己共分散) :**

$$\hat{\gamma}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T} (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu})$$

7. **Correlogram (コレログラム, or 標本自己相関関数) :**

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$$

8. **Lag Operator (ラグ作要素) :**

$$L^{\tau} y_t = y_{t-\tau}, \qquad \tau = 1, 2, \cdots$$

9. **Likelihood Function (尤度関数) —** Innovation Form :

The joint distribution of $y_1, y_2, \cdots, y_T$ is written as:

$$\begin{aligned}
f(y_1, , y_2, \cdots, y_T) &= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1}, \cdots, y_1) \\
&= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) f(y_{T-2}, \cdots, y_1) \\
&\ \vdots \\
&= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) \ \cdots \ f(y_2 | y_1) f(y_1) \\
&= f(y_1) \prod_{t=2}^{T} f(y_t | y_{t-1}, \cdots, y_1).
\end{aligned}$$

Therefore, the log-likelihood function is given by:

$$\log f(y_1, y_2, \cdots, y_T) = \log f(y_1) + \sum_{t=2}^{T} \log f(y_t | y_{t-1}, \cdots, y_1).$$

Under the normality assumption, $f(y_t|y_{t-1}, \cdots, y_1)$ is given by the normal distribution with conditional mean $\mathrm{E}(y_t|y_{t-1}, \cdots, y_1)$ and conditional variance $\mathrm{Var}(y_t|y_{t-1}, \cdots, y_1)$.

6.2 Time Series Models (時系列モデル)

Autoregressive Model (自己回帰モデル or AR モデル): AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

Moving Average Model (移動平均モデル or MA モデル): MA(q)

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$