4　系列相関：*DW* について

4.1　*DW* について

最小自乗法の仮定の一つに,「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」というものがあった。ダービン・ワトソン比 (DW) とは，誤差項の系列相関，すなわち，u_i と u_{i-1} との間の相関の有無を検定するために考案された。

$\Longrightarrow$ 時系列データのときのみ有効

$u_1, u_2, \cdots, u_n$ の系列について，それぞれの符号が，+ + + - - - - + + - - - + + のように，プラスが連続で続いた後で，マイナスが連続で続くというような場合，$u_1, u_2, \cdots, u_n$ は正の系列相関があると言う。また，+ - + - + - + - + のように交互にプラス，マイナスになる場合，$u_1, u_2, \cdots, u_n$ 負の系列相関があると言う。

特徴： $u_1, u_2, \cdots, u_i$ から u_{i+1} の符号が予想できる。$\Longrightarrow$ 「$u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」という仮定に反する。

すなわち，ダービン・ワトソン比とは，回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときに，H_0：$\rho = 0$, H_1：$\rho \neq 0$ の検定である。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

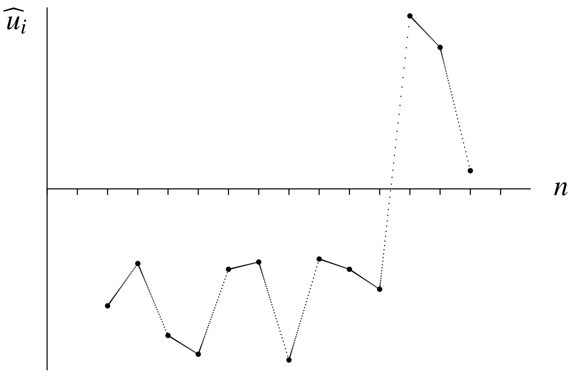

図 4： 正の系列相関

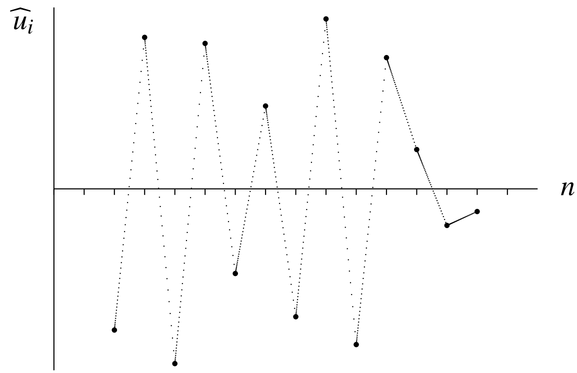

図 5： 負の系列相関

ダービン・ワトソン比の定義は次の通りである。

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n} \widehat{u}_i^2}$$

DW は近似的に，次のように表される。

$$\begin{aligned} DW &= \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n} \widehat{u}_i^2} = \frac{\sum_{i=2}^{n} \widehat{u}_i^2 - 2\sum_{i=2}^{n} \widehat{u_i}\widehat{u}_{i-1} + \sum_{i=2}^{n} \widehat{u}_{i-1}^2}{\sum_{i=1}^{n} \widehat{u}_i^2} \\ &= \frac{2\sum_{i=1}^{n} \widehat{u}_i^2 - (\widehat{u}_1^2 + \widehat{u}_n^2)}{\sum_{i=1}^{n} \widehat{u}_i^2} - 2\frac{\sum_{i=2}^{n} \widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=1}^{n} \widehat{u}_i^2} \approx 2(1 - \widehat{\rho}), \end{aligned}$$

以下の 2 つの近似が用いられる。

$$\begin{aligned} &\frac{\widehat{u}_1^2 + \widehat{u}_n^2}{\sum_{i=1}^{n} \widehat{u}_i^2} \approx 0, \\ &\frac{\sum_{i=2}^{n} \widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=1}^{n} \widehat{u}_i^2} = \frac{\sum_{i=2}^{n} \widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=2}^{n} \widehat{u}_{i-1}^2 + \widehat{u}_n^2} \approx \frac{\sum_{i=2}^{n} \widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=2}^{n} \widehat{u}_{i-1}^2} = \widehat{\rho}, \end{aligned}$$

すなわち，$\widehat{\rho}$ は $\widehat{u_i}$ と $\widehat{u}_{i-1}$ の回帰係数である。$u_i = \rho u_{i-1} + \epsilon_i$ において，u_i, u_{i-1} の代わりに $\widehat{u_i}, \widehat{u}_{i-1}$ に置き換えて，ρ の推定値 $\widehat{\rho}$ を求める。

1. DW の値が 2 前後のとき，系列相関なし ($\widehat{\rho} = 0$ のとき，$DW \approx 2$)。

2. DW が 2 より十分に小さいとき，正の系列相関と判定される。

3. DW が 2 より十分に大きいとき，負の系列相関と判定される。

正確な判定には，データ数 n とパラメータ数 k に依存する。表 1 を参照せよ。

k' は定数項を除くパラメータ数を表すものとする。

See `http://www.stanford.edu/~clint/bench/dwcrit.htm` for the DW table.

Table 1: ダービン・ワトソン統計量の 5 % 点の上限と下限

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$		k'
n	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.367	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	—	—	—	—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	—
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	0.09
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184	0.13
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073	0.17
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974	0.22
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885	0.26
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806	0.30
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735	0.34
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670	0.39
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613	0.43
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560	0.47
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513	0.50
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470	0.54
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431	0.57

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u_{i-1}})^2}{\sum_{i=1}^{n}\widehat{u_i^2}} \approx 2(1 - \widehat{\rho}) \longrightarrow 2(1 - \rho)$$

$-1 < \rho < 1$ なので (証明略)，近似的に $0 \leq DW \leq 4$ となる。

- $0 \leq DW \leq dl \quad \longrightarrow$ u_i に正の系列相関
- $dl \leq DW \leq du \quad \longrightarrow$ u_i に正の系列相関と判定できない
- $du \leq DW \leq 4 - du \quad \longrightarrow$ u_i に系列相関なし
- $4 - du \leq DW \leq 4 - dl \quad \longrightarrow$ u_i に負の系列相関と判定できない
- $4 - dl \leq DW \leq 4 \quad \longrightarrow$ u_i に負の系列相関

数値例：　今までと同じ数値例で，DW を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

$$
\begin{aligned}
DW &= \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \\
&= \frac{(-0.8-0.9)^2 + (0.9-0.6)^2 + (0.6-(-0.7))^2}{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2} = \frac{4.67}{2.30} = 2.03
\end{aligned}
$$

推定結果の表記方法：　回帰モデル：

$$Y_i = \alpha + \beta X_i + u_i,$$

の推定の結果，$\widehat{\alpha} = 0.3, \widehat{\beta} = 0.65, s_{\widehat{\alpha}} = \sqrt{10.0005} = 3.163, s_{\widehat{\beta}} = \sqrt{0.0575} = 0.240,$ $\dfrac{\widehat{\alpha}}{s_{\widehat{\alpha}}} = 0.095, \dfrac{\widehat{\beta}}{s_{\widehat{\beta}}} = 2.708, s^2 = 1.15$ (すなわち，$s = 1.07$), $R^2 = 0.786, \overline{R}^2 = 0.679,$ $DW = 2.03$ を得た。

これらをまとめて，

$$Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65}\,X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は t 値を表すものとする。

または，

$$Y_i = \underset{(3.163)}{0.3} + \underset{(0.240)}{0.65} X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は標準誤差を表すものとする。

のように書く。$s = \sqrt{1.15} = 1.07$ に注意。

4.2 系列相関のもとで回帰式の推定

回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \alpha(1-\rho) + \beta(X_i - \rho X_{i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_i^* = (X_i - \rho X_{i-1})$

を新たな変数として，

$$Y_i^* = \alpha' + \beta X_i^* + \epsilon_i,$$

に最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。ただし，$\alpha' = \alpha(1-\rho)$ の関係が成り立つことに注意。

より一般的に，回帰式が

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

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \beta_1(X_{1i} - \rho X_{1,i-1}) + \beta_2(X_{1i} - \rho X_{2,i-1}) + \cdots + \beta_k(X_{1i} - \rho X_{k,i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \rho X_{k,i-1})$ を新たな変数として，

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

最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。

ρ の求め方について (その 1): DW は近似的に $DW \approx 2(1 - \widehat{\rho})$ と表されるので，DW から ρ の推定値 $\widehat{\rho}$ を逆算して，

$Y_i^* = (Y_i - \widehat{\rho} Y_{i-1}),\ X_{1i}^* = (X_{1i} - \widehat{\rho} X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \widehat{\rho} X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \widehat{\rho} X_{k,i-1})$
を新たな変数として，

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i,$$

に最小二乗法を適用する。

ρ の求め方について (その 2):　収束計算によって求める。 $\longrightarrow$ コクラン・オーカット法

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

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\beta}_1, \cdots, \widehat{\beta}_k, \widehat{u}_i$ を得る。

2. $\widehat{u}_i = \rho \widehat{u}_{i-1} + \epsilon_i, \quad i = 2, 3, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\rho}$ を得る。

3. $\rho^{(m-1)} = \widehat{\rho}$ とおく。

4. $Y_i^* = (Y_i - \rho^{(m-1)} Y_{i-1})$, $X_{1i}^* = (X_{1i} - \rho^{(m-1)} X_{1,i-1})$, $X_{2i}^* = (X_{2i} - \rho^{(m-1)} X_{2,i-1})$, $\cdots$, $X_{ki}^* = (X_{ki} - \rho^{(m-1)} X_{k,i-1})$ を計算する。

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i, \quad i = 2, 3, \cdots, n$$

を最小二乗法で推定する。 $\longrightarrow \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ を得る。

5. $\widehat{u}_i = Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}, \quad i = 1, 2, \cdots, n$

を計算する。

6. ステップ 2 に戻り，$m = 1, 2, \cdots$ について繰り返す。

収束先を $\beta_1, \beta_2, \cdots, \beta_k, \rho$ の推定値とする。

5　不均一分散 (不等分散)

回帰式が

$$Y_i = \alpha + \beta X_i + u_i$$

の場合を考える。X_i が外生変数，Y_i は内生変数，u_i は互いに独立な同一の分布を持つ攪乱項 (最小二乗法に必要な仮定) とする。「独立な同一の分布」の意味は「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に平均ゼロ，分散 σ^2 の分布する」である。分散が時点に依存する場合，代表的には，分散が他の変数 (例えば，z_i) に依存する場合，すなわち，u_i の平均はゼロ，分散は $\sigma_*^2 z_i^2$ の場合は，最小二乗法の仮定に反する。そのため，単純には，$Y_i = \alpha + \beta X_i + u_i$ に最小二乗法を適用できない。以下のような修正が必要となる。

$$\frac{Y_i}{z_i} = \alpha\frac{1}{z_i} + \beta\frac{X_i}{z_i} + \frac{u_i}{z_i} = \alpha\frac{1}{z_i} + \beta\frac{X_i}{z_i} + u_i^*$$

このとき，新たな攪乱項 u_i^* は平均ゼロ，分散 σ_*^2 の分布となる (すなわち,「同

一の」分布)。

$$\mathrm{E}(u_i^*) = \mathrm{E}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)\mathrm{E}(u_i) = 0$$

u_i の仮定 $\mathrm{E}(u_i) = 0$ が使われている。

$$\mathrm{V}(u_i^*) = \mathrm{V}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)^2 \mathrm{V}(u_i) = \sigma_*^2$$

u_i の仮定 $\mathrm{V}(u_i) = \sigma_*^2 z_i^2$ が最後に使われている。

よって，$\dfrac{Y_i}{z_i}, \dfrac{1}{z_i}, \dfrac{X_i}{z_i}$ を新たな変数として，最小二乗法を適用することができる。

不均一分散の検定について

$$\widehat{u_i^2} = \gamma z_i + \epsilon_i$$

を推定し，γ の推定値 $\widehat{\gamma}$ の有意性の検定を行う (通常の t 検定)。

z_i は回帰式に含まれる変数でもよい。例えば，u_i の平均はゼロ，分散は $\sigma_*^2 X_i^2$ の

場合，各変数を X_i で割って，

$$\frac{Y_i}{X_i} = \alpha\frac{1}{X_i} + \beta + \frac{u_i}{X_i} = \alpha\frac{1}{X_i} + \beta + u_i^*$$

を推定すればよい。β は定数項として推定されるが，意味は限界係数 (すなわち，傾き) と同じなので注意すること。

6　推定量の求め方

6.1　最小二乗法

・n 個のデータ (実現値)： $x_1, x_2, \cdots, x_n$

・背後に対応する確率変数を仮定： $X_1, X_2, \cdots, X_n$

・$\mathrm{E}(X_i) = \mu$，$\mathrm{V}(X_i) = \sigma^2$ を仮定

母数 (μ, σ^2) を推定する。

観測データ $x_1, x_2, \cdots, x_n$ をもとにして，μ の最小二乗推定値を求める。

$$\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2$$

μ の解を $\widehat{\mu}$ とすると，

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

となり，$\widehat{\mu} \equiv \overline{x}$ を得る。

すなわち，

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

を μ について解く。

μ の最小二乗推定量はデータ x_i を対応する確率変数 X_i で置き換えて，

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

となり，$\widehat{\mu} \equiv \overline{X}$ を得る ($\widehat{\mu}$ について，推定値と推定量は同じ記号を使っている)。

以上を回帰分析に応用すると，

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

を解くことになる。

すなわち，

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

の連立方程式を α, β について解く。

6.2　最尤法

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 $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \cdots, x_n)$ 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，$\widehat{\theta} = \widehat{\theta}(X_1, X_2, \cdots, X_n)$ 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，

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

ただし，

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

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

θ がベクトル $(k \times 1)$ の場合，n が大きいとき，

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

ただし，

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

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，

(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］$\mathrm{N}(\mu, \sigma^2)$ の密度関数は,

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

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,

$$\begin{aligned} f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) &\equiv \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Bigl(-\frac{1}{2\sigma^2}(x_i-\mu)^2\Bigr) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\Bigr) \end{aligned}$$

となる。

(1) σ^2 が既知のとき，尤度関数 $l(\mu)$ は,

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

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,

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

となり,

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

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

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

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,

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

となる。

$\widehat{\mu}$ の分散を求めるために,

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

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

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

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

と計算される。

最尤推定量の性質から，n が大きいとき，

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

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，$\widehat{\mu} \sim N(\mu, \sigma_\mu^2)$ 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 $\widehat{\mu}, \widehat{\sigma}^2$ とすると，最尤推定値は,

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

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

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,

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

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

となる。

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

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

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

ただし,

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

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

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

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

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

よって，

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

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

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

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 $\widehat{p}$ は，

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

となる。

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

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

となる。

$\widehat{p}$ の分布を求める。

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

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

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

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

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

したがって，

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

を得る。

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,

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

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 $\widehat{\lambda}$ は,

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

となる。

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

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

となる。

$\widehat{\lambda}$ の分布を求める。

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

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

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

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

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

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

したがって,

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

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 $\widehat{\lambda}$ は,

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

となる。

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

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

となる。

$\widehat{\lambda}$ の分布を求める。

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

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

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

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

したがって，

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

を得る。

6.2.1　変数変換

確率変数 X の密度関数を $f(x)$，分布関数を $F(x) \equiv \mathrm{P}(X < x)$ とする。$Y = aX + b$ とするとき，Y の密度関数 $g(y)$ を求める。

Y の分布関数を $G(y)$ として，次のように変形できる。

$$
\begin{aligned}
G(y) &= \mathrm{P}(Y < y) = \mathrm{P}(aX + b < y) \\
&= \begin{cases} \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ \mathrm{P}\left(X > \dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases} \\
&= \begin{cases} \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ 1 - \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases}
\end{aligned}
$$

$$= \begin{cases} F\left(\dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ 1 - F\left(\dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases}$$

分布関数と密度関数との関係は,

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x) \qquad \frac{\mathrm{d}G(x)}{\mathrm{d}x} = g(x)$$

であるので, Y の密度関数は,

$$\begin{aligned} g(y) &= \frac{\mathrm{d}G(y)}{\mathrm{d}y} \\ &= \begin{cases} \dfrac{1}{a} f\left(\dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ -\dfrac{1}{a} f\left(\dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases} \\ &= \left|\frac{1}{a}\right| f\left(\frac{y-b}{a}\right) \end{aligned}$$

と表される。

一般に，確率変数 X の密度関数を $f(x)$ とする。単調変換 $X = h(Y)$ とするとき，Y の密度関数 $g(y)$ は，

$$g(y) = |h'(y)| f(h(y))$$

となる。

6.2.2　回帰分析への応用

回帰モデル

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1, 2, \cdots, n$$

$u_1, u_2, \cdots, u_n$ は互いに独立で，すべての i について $u_i \sim \mathrm{N}(0, \sigma^2)$ を仮定する。

u_i の密度関数は，

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Bigl(-\frac{1}{2\sigma^2} u_i^2\Bigr)$$

となる。

Y_i の密度関数 $g(Y_i)$ は，

$$g(Y_i) = |h'(Y_i)| f(h(Y_i))$$

によって求められる。

この場合，$h(Y_i) = Y_i - \alpha - \beta X_i$ なので，$h'(Y_i) = 1$ となる。

したがって，Y_i の密度関数は，

$$\begin{aligned} g(Y_i) &= |h'(Y_i)| f(h(Y_i)) \\ &= f(h(Y_i)) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Bigl(-\frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2\Bigr) \end{aligned}$$

となる。

$u_1, u_2, \cdots, u_n$ は互いに独立であれば，$Y_1, Y_2, \cdots, Y_n$ も互いに独立になるので，Y_1,

$Y_2, \cdots, Y_n$ の結合密度関数は，

$$
\begin{aligned}
g(Y_1, Y_2, \cdots, Y_n) &= \prod_{i=1}^{n} g(Y_i) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2\Big)
\end{aligned}
$$

となる。これは α, β, σ^2 の関数となっている。

よって，尤度関数は，

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

となる。

対数尤度関数は，

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

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，

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

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$$

σ^2 の最尤推定量は,

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

となり, s^2 とは異なる。

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

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

ただし,

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

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。

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

$$= \begin{pmatrix} \dfrac{1}{\sigma^2}(Y_i - \alpha - \beta X_i) \\ \dfrac{1}{\sigma^2}X_i(Y_i - \alpha - \beta X_i) \\ -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(Y_i - \alpha - \beta X_i)^2 \end{pmatrix}$$

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

ただし，$u_i = Y_i - \alpha - \beta X_i$

期待値をとると，

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

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

となる。

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

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

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

したがって，

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

となる。

$\longrightarrow$ 最小二乗推定量の分布と同じ。