7.4　説明変数：確率変数のケース

本節では，簡単化のために，単回帰 $Y_i = \alpha + \beta X_i + u_i$ を考える。

(1) X_i は非確率変数 $\longrightarrow$ 今までの最小二乗推定量

(2) X_i は確率変数

(2a) X_i と u_i は相関なし $\longrightarrow$ $\mathbf{Cov}(X_i, u_i) = 0$（**7.4.1** 節）

(2b) X_i は u_i は相関あり $\longrightarrow$ $\mathbf{Cov}(X_i, u_i) \neq 0$（**7.4.2** 節）

7.4.1　説明変数と誤差項に相関がない場合

最小二乗法による推定量：

$$\hat{\beta} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \beta + \sum \omega_i u_i$$

ただし， $\omega_i = \dfrac{X_i - \overline{X}}{\sum (X_j - \overline{X})^2}$ **とする。**

$X = (X_1,\ X_2,\ \cdots,\ X_n)$ **とする。**

(* 復習) 2 つの確率変数（X，Y）の独立について：

X と Y の同時密度関数 $f_{xy}(x, y)$

X の周辺密度関数 $f_x(x)$

Y の周辺密度関数 $f_y(y)$

Y を与えたもとで X の条件付き密度関数 $f_{x|y}(x|y)$

- $f_{xy}(x, y) = f_{x|y}(x|y)f_y(y)$ **は必ず成り立つ。**
- $f_{xy}(x, y) = f_x(x)f_y(y) \quad \Longleftrightarrow \quad X$ **と** Y **は独立**
- $f_{x|y}(x|y) = f_x(x) \quad \Longleftrightarrow \quad X$ **と** Y **は独立**

(* 復習) 2 つの確率変数（X，Y）の独立について（その 2）：

X と Y が独立のとき，

- $\mathbf{Cov}(X,Y)=0$
- $\mathbf{E}(X|Y)=\mathbf{E}(X)$，$\mathbf{V}(X|Y)=\mathbf{V}(X)$

となる。

条件付き期待値を取ると，

$$\mathbf{E}(\hat{\beta}|X)=\mathbf{E}(\beta+\sum_i\omega_iu_i|X)=\beta+\sum_i\omega_i\mathbf{E}(u_i|X)=\beta$$

となり，$\hat{\beta}$ は β の不偏推定量となる。

ω_i は X_1，X_2，$\cdots$，X_n の関数となっている。

$X = (X_1, X_2, \cdots, X_n)$ と u_i に相関がない場合，すなわち，$j = 1, 2, \cdots, n$ について $\mathbf{Cov}(X_j, u_i) = 0$ の場合，$\mathbf{E}(u_i|X) = \mathbf{E}(u_i) = 0$ となる。

条件付き分散については，

$$\mathbf{V}(\hat{\beta}|X) = \mathbf{V}(\beta + \sum_i \omega_i u_i|X) = \mathbf{V}(\sum_i \omega_i u_i|X) = \sum_i \omega_i^2 \mathbf{V}(u_i|X) = \sigma^2 \sum_i \omega_i^2 = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

となる。

すなわち，説明変数が確率変数であっても，誤差項と相関がなければ，何も変更せずに，最小二乗法を適用することができる。

7.4.2　説明変数と誤差項に相関がある場合

X と u_i に相関がある場合，$\mathbf{E}(u_i|X) \neq \mathbf{E}(u_i) = 0$ となるので，

$$\mathbf{E}(\hat{\beta}|X) = \mathbf{E}(\beta + \sum_i \omega_i u_i|X) = \beta + \sum_i \omega_i \mathbf{E}(u_i|X) \neq \beta$$

となる。

したがって，$\hat{\beta}$ は β の不偏推定量とはならない。

$\hat{\beta}$ は β の一致推定量かどうか？

$$\hat{\beta} = \beta + \sum_i \omega_i u_i = \beta + \frac{\sum_i (X_i - \overline{X}) u_i}{\sum_j (X_j - \overline{X})^2}$$

$$= \beta + \frac{\frac{1}{n}\sum_i (X_i - \overline{X})u_i}{\frac{1}{n}\sum_j (X_j - \overline{X})^2} \longrightarrow \beta + \frac{M_{xu}}{M_{xx}} \neq \beta$$

ただし，$n \longrightarrow \infty$ のとき，$\frac{1}{n}\sum_i (X_i - \overline{X})u_i \longrightarrow M_{xu} \neq 0$，$\frac{1}{n}\sum_i (X_i - \overline{X})^2 \longrightarrow M_{xx}$ とする。$n \longrightarrow \infty$ のときは，分子・分母は別々に計算することができる（証明略）。

M_{xu} は，$n \longrightarrow \infty$ のとき，X_i と u_i の共分散に相当する。M_{xx} は，$n \longrightarrow \infty$ のとき，X_i の分散に相当する。

以上から，$\hat{\beta}$ は β の不偏推定量でも一致推定量でもない。

$\hat{\alpha}$ も同様に不偏推定量でも一致推定量でもない。

なぜなら，$\hat{\alpha}$ は，

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = \alpha + \sum_i \lambda_i u_i$$

と書き換えられる（5.3.2 節参照）。

ただし，$\lambda_i = \dfrac{1}{n} - \overline{X}\omega_i$ とする。

λ_i は X の関数である。

● 例：X に観測誤差（measurement error）が含まれる場合：　真のモデルを

$$Y_i^* = \alpha + \beta X_i^*$$

とする。$(Y_i^*$，$X_i^*)$ は非確率変数とする。

しかし，(Y_i^*, X_i^*) は観測されず，代わりに，(Y_i, X_i) が観測されるものとする。

(Y_i^*, X_i^*) と (Y_i, X_i) との関係は以下の通りとする。

$$Y_i = Y_i^* + u_i, \qquad X_i = X_i^* + v_i$$

u_i，v_i は観測誤差と呼ばれるもので，

$$\mathbf{E}(u_i) = 0, \qquad \mathbf{V}(u_i) = \sigma_u^2$$

$$\mathbf{E}(v_i) = 0, \qquad \mathbf{V}(v_i) = \sigma_v^2$$

を仮定する。

さらに，u_i，v_i は互いに独立と仮定する。

すなわち，$i \neq j$ となるすべての i，j について $\mathbf{Cov}(u_i, u_j) = \mathbf{Cov}(v_i, v_j) = 0$，かつ，すべての i，j について $\mathbf{Cov}(u_i, v_j) = 0$ とする。

$Y_i^* = \alpha + \beta X_i^*$ に $Y_i = Y_i^* + u_i$，$X_i = X_i^* + v_i$ を代入する。

$$Y_i = \alpha + \beta X_i + (u_i - \beta v_i)$$

観測されるのは $(Y_i$，$X_i)$ なので，$(u_i - \beta v_i)$ を誤差項として，最小二乗法で $\hat{\beta}$ を求める。

まずは，X_i と $u_i - \beta v_i$ の共分散を求める（共分散がゼロかどうかを確認する）。

(* 復習) 共分散について：

2 つの確率変数 (X, Y) を考える。

$\mathbf{E}(X) = \mu_x$，$\mathbf{E}(Y) = \mu_y$ **とする。**

共分散の定義は， $\mathbf{Cov}(X, Y) = \mathbf{E}\Big((X - \mu_x)(Y - \mu_y)\Big)$

書き換えると，

$$\mathbf{Cov}(X, Y) = \mathbf{E}(XY - X\mu_y - \mu_x Y + \mu_x\mu_y) = \mathbf{E}(XY) - \mathbf{E}(X)\mu_y - \mu_x\mathbf{E}(Y) + \mu_x\mu_y = \mathbf{E}(XY) - \mu_x\mu_y$$

となる。

この場合，

$$\begin{aligned}\mathbf{Cov}(X_i,\ u_i-\beta v_i) &= \mathbf{Cov}(X_i^*+v_i,\ u_i-\beta v_i)\\ &= \mathbf{E}\Big((X_i^*+v_i)(u_i-\beta v_i)\Big)-\mathbf{E}(X_i^*+v_i)\mathbf{E}(u_i-\beta v_i)\\ &= \mathbf{E}(X_i^*u_i+v_iu_i-X_i^*\beta v_i-\beta v_i^2)\\ &= \mathbf{E}(X_i^*u_i)+\mathbf{E}(v_iu_i)-\mathbf{E}(X_i^*\beta v_i)-\mathbf{E}(\beta v_i^2)\\ &= X_i^*\mathbf{E}(u_i)+\mathbf{E}(v_iu_i)-X_i^*\beta\mathbf{E}(v_i)-\beta\mathbf{E}(v_i^2)\\ &= -\beta\sigma_v^2\neq 0\end{aligned}$$

となる。

したがって，観測できる $(Y_i$，$X_i)$ を用いて，β の最小二乗推定量 $\hat{\beta}$ は不偏推定量にはならない。

特に，

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} = \beta + \sum \omega_i(u_i - \beta v_i) \\
&= \beta + \frac{\sum(X_i - \overline{X})(u_i - \beta v_i)}{\sum(X_i - \overline{X})^2} \\
&= \beta + \frac{\frac{1}{n}\sum(X_i - \overline{X})(u_i - \beta v_i)}{\frac{1}{n}\sum(X_i - \overline{X})^2}
\end{aligned}
$$

と書き換えられ，右辺第 2 項の分母は X_i の分散に対応し，分子は X_i と $(u_i - \beta v_i)$ との共分散 $-\beta\sigma_v^2$ に対応する。

したがって，$n \longrightarrow \infty$ のとき，

$$
\hat{\beta} \longrightarrow \beta + \frac{-\beta\sigma_v^2}{M_{xx}}
$$

となる。右辺第 2 項の分母は必ず正，分子は β が正（負）の場合は負（正）となる。

すなわち，

- $\beta > 0$ のとき，$\hat{\beta} \longrightarrow \beta - \dfrac{\beta\sigma_v^2}{M_{xx}} < \beta$
- $\beta < 0$ のとき，$\hat{\beta} \longrightarrow \beta - \dfrac{\beta\sigma_v^2}{M_{xx}} > \beta$

となる。

● X と u_i に相関がある場合の対処法： $Y_i = \alpha + \beta X_i + u_i$ について，$\mathbf{Cov}(X_i, u_i) \neq 0$ のときを考える。

$$\hat\beta = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2} = \beta + \frac{\sum_i (X_i - \overline{X}) u_i}{\sum_j (X_j - \overline{X})^2}$$

$$= \beta + \frac{\frac{1}{n}\sum_i (X_i - \overline{X}) u_i}{\frac{1}{n}\sum_j (X_j - \overline{X})^2} \longrightarrow \beta + \frac{M_{xu}}{M_{xx}} \neq \beta$$

右辺第 2 項の分母は X_i の分散に相当し，分子は X_i と u_i の共分散に相当する（$n \longrightarrow \infty$ のときは，分子・分母を別々に計算することができる)。

$\mathbf{Cov}(X_i, u_i) \neq 0$ が問題となって，$\mathbf{E}(\hat\beta) \neq \beta$ となる。

よって，第 2 項の分子がゼロになるような修正を加えればよい。

$\mathbf{Cov}(Z_i, u_i) = 0$ となる Z_i が存在するとする。

このとき，下記のような推定量

$$\widetilde{\beta} = \frac{\sum_i (Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})}$$

を考えてみよう。ただし，$\overline{Z} = \dfrac{1}{n}\displaystyle\sum_i Z_i$ とする。

(*) Z_i を X_i で置き換えると，$\widetilde{\beta}$ は最小二乗推定量 $\hat{\beta} = \dfrac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}$ に等しくなる。

$\widetilde{\beta}$ を変形していく。

$$
\begin{aligned}
\widetilde{\beta} &= \frac{\sum_i (Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \frac{\sum_i (Z_i - \overline{Z})Y_i - \overline{Y}\sum_i (Z_i - \overline{Z})}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \frac{\sum_i (Z_i - \overline{Z})Y_i}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \qquad \left(= \sum_i \omega_i^* Y_i, \quad \text{ただし，}\ \omega_i^* = \frac{Z_i - \overline{Z}}{\sum_j (Z_j - \overline{Z})(X_j - \overline{X})}\right) \\
&= \frac{\sum_i (Z_i - \overline{Z})(\alpha + \beta X_i + u_i)}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \frac{\alpha \sum_i (Z_i - \overline{Z}) + \beta \sum_i (Z_i - \overline{Z})X_i + \sum_i (Z_i - \overline{Z})u_i}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \\
&= \beta + \frac{\sum_i (Z_i - \overline{Z})u_i}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \qquad \left(= \beta + \sum_i \omega_i^* u_i\right) \\
&= \beta + \frac{\frac{1}{n}\sum_i (Z_i - \overline{Z})u_i}{\frac{1}{n}\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \longrightarrow \beta + \frac{0}{M_{zx}} = \beta
\end{aligned}
$$

2 行目の右辺の分子の第 2 項目は $\sum_i (Z_i - \overline{Z}) = \sum_i Z_i - n\overline{Z} = 0$ に注意。

3 行目では，$\widetilde{\beta} = \dfrac{\sum_i (Z_i - \overline{Z}) Y_i}{\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} = \sum_i \omega_i^* Y_i$ と書き換えることができ，$\widetilde{\beta}$ も Y_i の線形推定量と言える。

ただし，$\omega_i^* = \dfrac{Z_i - \overline{Z}}{\sum_j (Z_j - \overline{Z})(X_j - \overline{X})}$ である（分母の添字を i から j に変更）。

4 行目の右辺分子の第 1 項目はゼロ，第 2 項目は $\sum_i (Z_i - \overline{Z}) X_i = \sum_i (Z_i - \overline{Z}) X_i - \sum_i (Z_i - \overline{Z}) \overline{X} = \sum_i (Z_i - \overline{Z})(X_i - \overline{X})$ となるので分母と同じになる。このようにして，5 行目が得られる。

6 行目右辺第 2 項の分子は Z_i と u_i の共分散に対応し，分母は Z_i と X_i の共分散に対応し，n を大きくするとそれぞれゼロ，M_{zx} に収束するものとする。

すなわち，$\widetilde{\beta}$ は β の一致推定量となる。

n が大きければ，$\mathbf{E}(\omega_i^* u_i) \longrightarrow 0$ となる（分子・分母を別々に計算することができる)。

しかし，一般的には，$\mathbf{E}(\omega_i^* u_i) \neq 0$ なので（ω_i^* の分母は X_i に依存していて，X_i と u_i は共分散がゼロでないと仮定），$\mathbf{E}(\widetilde{\beta}) \neq \beta$ となり，$\widetilde{\beta}$ は不偏推定量にはならない。

Z_i を操作変数（**instrumental variable**）と呼ぶ。操作変数を用いた推定方法を操作変数法という。

Z_i の選択について，**(i)** Z_i と u_i は相関がない，**(ii)** Z_i と X_i は強い相関がある，という **2** つの条件が必要になる。

(ii) については，Z_i はもともと X_i の代わりに使うものなので，X_i と相関の強い Z_i が望ましい。

● Z_i の選択について（その 1）：　i が時間を表す場合（時系列データの場合），$Z_i = X_{i-1}$ を用いることが可能である。

$\mathbf{Cov}(X_i, u_i) \neq 0$ としても，X を一期ずらして $\mathbf{Cov}(X_{i-1}, u_i) = 0$ となるのは不自然ではない。

● Z_i の選択について（その 2）：　X_i の予測値 $\hat{X}_i$ を X_i の代わりに用いる。

e_i を誤差項として，

$$X_i = \gamma_0 + \gamma_1 W_{1i} + \gamma_2 W_{2i} + \cdots + \gamma_m W_{mi} + e_i$$

を最小二乗法で推定して，$\hat{\gamma}_0$，$\hat{\gamma}_1$，$\cdots$，$\hat{\gamma}_m$ を求める。

ただし，W_{1i}，W_{2i}，$\cdots$，W_{mi} は u_i と相関のない変数でなければならない。

W_{1i}，W_{2i}，$\cdots$，W_{mi} には，X_{i-1}，X_{i-2}，$\cdots$ のように X_i のラグ変数を用いてもよい。理由は，前述の通りで，X_i と u_i に相関かあったとしても，X_i のラグ変数 X_{i-1}，X_{i-2}，$\cdots$ と u_i とに相関があるとは考えにくいからである。

X_i の予測値 $\hat{X}_i$ を求める。

$$\hat{X}_i = \hat{\gamma}_0 + \hat{\gamma}_1 W_{1i} + \hat{\gamma}_2 W_{2i} + \cdots + \hat{\gamma}_m W_{mi}$$

を Z_i として用いる。

Z_i は u_i と相関のない変数でなければならない。

$\hat{\gamma}_j$ は γ_j の一致推定量なので，$n \longrightarrow \infty$ のとき，

$$\hat{X}_i = \hat{\gamma}_0 + \hat{\gamma}_1 W_{1i} + \hat{\gamma}_2 W_{2i} + \cdots + \hat{\gamma}_m W_{mi} \longrightarrow \gamma_0 + \gamma_1 W_{1i} + \gamma_2 W_{2i} + \cdots + \gamma_m W_{mi}$$

となる。

操作変数法による推定量$\widetilde{\beta}$がβの一致推定量になる理由は，

$$\widetilde{\beta} = \beta + \frac{\frac{1}{n}\sum_i (Z_i - \overline{Z})u_i}{\frac{1}{n}\sum_i (Z_i - \overline{Z})(X_i - \overline{X})} \longrightarrow \beta + \frac{0}{M_{zx}} = \beta$$

から，操作変数Z_iと誤差項u_iとの相関がゼロという条件（2 項目の分子）が重要なポイントとなっている。

Z_iに$\hat{X}_i$を用いると，$n \longrightarrow \infty$のとき，$\hat{X}_i \longrightarrow \gamma_0 + \gamma_1 W_{1i} + \gamma_2 W_{2i} + \cdots + \gamma_m W_{mi}$となることから，$W_{1i}$，$W_{2i}$，$\cdots$，$W_{mi}$が$u_i$と相関がなければ，$\frac{1}{n}\sum_i (\hat{X}_i - \overline{X})u_i \longrightarrow 0$となる（$\overline{\hat{X}} = \overline{X}$に注意）。

この方法は，二段階最小二乗法（two-stage least squares method）と呼ばれる。1 段階目

で $\hat{X}_i$ を求める。2 段階目で $\widetilde{\beta}$ を得る。