例：

$X_1, X_2, \cdots, X_n$ は互いに独立で，すべて同一のポアソン分布 (すなわち，平均 λ ですべて同一の分布) に従うものとする。λ の最尤推定量を求める。

ポアソン分布の確率関数は,

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

なので，尤度関数は,

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

対数尤度関数は,

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

となる。

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

これを解いて，λ の最尤推定量 $\widehat{\lambda}$ は，

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

となる。

$\widehat{\lambda}$ は，λ の不偏推定量，有効推定量，十分推定量，一致推定量である。

証明：

X がパラメータ λ のポアソン分布に従うとき，

$$\mathrm{E}(X) = \mathrm{V}(X) = \lambda$$

となる。

不偏性：

$$\mathrm{E}(\widehat{\lambda}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$

有効性：

$$\mathrm{V}(\widehat{\lambda}) = \mathrm{V}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \lambda = \frac{\lambda}{n}$$

$$\begin{aligned}\frac{1}{n\mathrm{E}\left[\left(\frac{\partial \log f(X;\lambda)}{\partial\lambda}\right)^2\right]} &= \frac{1}{n\mathrm{E}\left[\left(\frac{\partial (X\log\lambda - \lambda - \log X!)}{\partial\lambda}\right)^2\right]} = \frac{1}{n\mathrm{E}\left[\left(\frac{X}{\lambda} - 1\right)^2\right]} \\ &= \frac{\lambda^2}{n\mathrm{E}[(X-\lambda)^2]} = \frac{\lambda^2}{n\mathrm{V}(X)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n}\end{aligned}$$

したがって，

$$\mathrm{V}(\widehat{\lambda}) = \frac{1}{n\mathrm{E}\left[\left(\frac{\partial \log f(X;\lambda)}{\partial\lambda}\right)^2\right]}$$

となり，$\mathrm{V}(\widehat{\lambda})$ は，クラメール・ラオの下限に一致する。よって，$\widehat{\lambda}$ は有効推定量である。

十分性：

$$\prod_{i=1}^{n} f(x_i;\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} = \frac{\lambda^{n\overline{x}} e^{-n\lambda}}{(n\overline{x})!} \frac{(n\overline{x})!}{\prod_{i=1}^{n} x_i!} = g(\overline{x};\lambda)\, h(x_1, x_2, \cdots, x_n)$$

と分解できる。

一致性：

$$\mathrm{E}(\overline{X}) = \lambda, \qquad \mathrm{V}(\overline{X}) = \frac{\lambda}{n}$$

なので，チェビシェフの不等式に当てはめる。

$$P(|\overline{X} - \lambda| > \epsilon) < \frac{\lambda}{n\epsilon^2} \longrightarrow \infty$$

したがって，一致性も成り立つ。

6.1 最尤法の例：AR(1) モデル

$y_t = \phi y_{t-1} + \epsilon_t,\ \ \epsilon_t \sim N(0, \sigma^2)$

1. Mean of y_t given $y_{t-1}, y_{t-2}, \cdots$

$$\mathrm{E}(y_t | y_{t-1}, y_{t-2}, \cdots) = \phi y_{t-1}$$

2. Variance of y_t given $y_{t-1}, y_{t-2}, \cdots$

$$\mathrm{V}(y_t | y_{t-1}, y_{t-2}, \cdots) = \sigma^2$$

3. Thus, $y_t | y_{t-1}, y_{t-2}, \cdots \sim N(0, \sigma^2)$. $\Longrightarrow$ Conditional distribution of y_t given $y_{t-1}, y_{t-2}, \cdots$

4. The stationarity condition is: the solution of $\phi(x) = 1 - \phi x = 0$, i.e., $x = 1/\phi$, is greater than one in absolute value, or equivalently, $|\phi| < 1$.

5. Rewriting the AR(1) model,

$$y_t = \phi y_{t-1} + \epsilon_t$$

$$
\begin{aligned}
&= \phi^2 y_{t-2} + \epsilon_t + \phi\epsilon_{t-1} \\
&= \phi^3 y_{t-3} + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} \\
&\ \ \vdots \\
&= \phi^s y_{t-s} + \epsilon_t + \phi\epsilon_{t-1} + \cdots + \phi^{s-1}\epsilon_{t-s+1}.
\end{aligned}
$$

As s is large, ϕ^s approaches zero. $\Longrightarrow$ Stationarity condition

6. For stationarity, $y_t = \phi y_{t-1} + \epsilon_t$ is rewritten as:

$$
y_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots
$$

7. Mean of y_t

$$
\begin{aligned}
\mathrm{E}(y_t) &= \mathrm{E}(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots) \\
&= \mathrm{E}(\epsilon_t) + \phi\mathrm{E}(\epsilon_{t-1}) + \phi^2\mathrm{E}(\epsilon_{t-2}) + \cdots = 0
\end{aligned}
$$

8. Variance of y_t

$$\begin{aligned} V(y_t) &= V(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots) \\ &= V(\epsilon_t) + V(\phi\epsilon_{t-1}) + V(\phi^2\epsilon_{t-2}) + \cdots \\ &= \sigma^2(1 + \phi^2 + \phi^4 + \cdots) = \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

9. Thus, $y_t \sim N\left(0, \dfrac{\sigma^2}{1-\rho^2}\right)$. $\implies$ Unconditional distribution of y_t

10. Estimation of AR(1) model:

    (a) Log-likelihood function

$$\begin{aligned} \log f(y_T, \cdots, y_1) &= \log f(y_1) + \sum_{t=1}^{T} \log f(y_t|y_{t-1}, \cdots, y_1) \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{1}{\sigma^2/(1-\phi^2)}y_1^2 \end{aligned}$$

$$-\frac{T-1}{2}\log(2\pi)-\frac{T-1}{2}\log(\sigma^2)-\frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t-\phi y_{t-1})^2$$

$$=-\frac{T}{2}\log(2\pi)-\frac{T}{2}\log(\sigma^2)-\frac{1}{2}\log\left(\frac{1}{1-\phi^2}\right)$$

$$-\frac{1}{2\sigma^2/(1-\phi^2)}y_1^2-\frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t-\phi y_{t-1})^2$$

Note as follows:

$$f(y_1)=\frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}}\exp\left(-\frac{1}{2\sigma^2/(1-\phi^2)}y_1^2\right)$$

$$f(y_t|y_{t-1},\cdots,y_1)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y_t-\phi y_{t-1})^2\right)$$

$$\frac{\partial\log f(y_T,\cdots,y_1)}{\partial\sigma^2}=-\frac{T}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4/(1-\phi^2)}y_1^2+\frac{1}{2\sigma^4}\sum_{t=2}^{T}(y_t-\phi y_{t-1})^2=0$$

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \phi} = -\frac{\phi}{1-\phi^2} + \frac{\phi}{\sigma^2} y_1^2 + \frac{1}{\sigma^2} \sum_{t=2}^{T} (y_t - \phi y_{t-1}) y_{t-1} = 0$$

The MLE of ϕ and σ^2 satisfies the above two equation.

6.2 最尤法の例：系列相関のもとで回帰式の推定：その 2

$$y_t = X_t\beta + u_t, \qquad u_t = \rho u_{t-1} + \epsilon, \qquad \epsilon_t \sim N(0, \sigma^2)$$

Log of distribution function of u_t

$$\begin{aligned}
\log f(u_T, \cdots, u_1) &= \log f(u_1) + \sum_{t=1}^{T} \log f(u_t | u_{t-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{\sigma^2/(1-\rho^2)}u_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}(u_t - \rho u_{t-1})^2 \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) \\
&\quad -\frac{1}{2\sigma^2/(1-\rho^2)}u_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(u_t - \rho u_{t-1})^2
\end{aligned}$$

Log of distribution function of y_t

$$
\begin{aligned}
&\log f(y_T, \cdots, y_1) \\
&= \log f(y_1) + \sum_{t=1}^{T} \log f(y_t | y_{t-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{\sigma^2/(1-\rho^2)}(y_1 - X_1\beta)^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}\Big((y_t - X_t\beta) - \rho(y_{t-1} - X_{t-1}\beta)\Big)^2
\end{aligned}
$$

$$
= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t^* - X_t^*\beta)^2,
$$

where

$$
y_t^* = \begin{cases} \sqrt{1-\rho^2}\, y_t, & \text{for } t = 1, \\ y_t - \rho y_{t-1}, & \text{for } t = 2, 3, \cdots, T, \end{cases} \qquad
X_t^* = \begin{cases} \sqrt{1-\rho^2}\, X_t, & \text{for } t = 1, \\ X_t - \rho X_{t-1}, & \text{for } t = 2, 3, \cdots, T, \end{cases}
$$

$\log f(y_T, \cdots, y_1)$ is maximized with respect to β, ρ and σ^2.

推定例： OLS, AR(1), AR(1)+X

StataSE をクリック

● データの編集

「Data」「Data Editor」を選択

Excel からデータのコピー
123,456 という形式でなく，123456 のようにコンマのない形式に設定すること。
方法：「書式」「セル」のところで「表示形式」のタブの「標準」を選択
データ名は var1，var2，var3，... となるので，出来れば変更

● command の欄にコマンドを入力

例えば，Y=$\alpha+\beta$ X+γ Z で，α，β，γ を推定するとき，
「reg Y X Z」リターン
とタイプする。　結果は results の欄に出力

Y，X，Z が時系列データのとき，

「gen t=_n」リターン
「tsset t」リターン
として，時系列データを扱っているということを宣言する。　t は他の名前でも構わない。
そして，
「reg Y X Z」リターン
とする。
「dwstat」リターン
とすると，ダービンワトソン比が出力される。

グラフについて：
「scatter Y X」リターン
とすると，横軸 X，縦軸 Y のグラフ。
「line Y X time」リターン
とすると，横軸 time，縦軸 X と Y のグラフ。

● 参考書

筒井淳也、秋吉美都、水落正明、 福田亘孝著
『Stata で計量経済学入門』(2007 年 3 月) ミネルヴァ書房 \2,940

● データ： 山本拓 (1995)『計量経済学』の数値例

t	x	y
1	10	6
2	12	9
3	14	10
4	16	10

● 出力結果

```
. gen t=_n

. tsset t

. reg y x

      Source |       SS       df       MS              Number of obs =       4
-------------+------------------------------           F(  1,     2) =    7.35
       Model |      8.45       1        8.45           Prob > F      =  0.1134
    Residual |       2.3       2        1.15           R-squared     =  0.7860
-------------+------------------------------           Adj R-squared =  0.6791
       Total |     10.75       3  3.58333333           Root MSE      =  1.0724

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |        .65   .2397916     2.71   0.113    -.3817399     1.68174
       _cons |         .3   3.163068     0.09   0.933    -13.30958    13.90958
------------------------------------------------------------------------------

. arima y, ar(1) nocons

(setting optimization to BHHH)
Iteration 0:   log likelihood = -10.213007
```

```
Iteration 1:   log likelihood = -9.8219683
Iteration 2:   log likelihood = -9.7761938
Iteration 3:   log likelihood = -9.6562972
Iteration 4:   log likelihood = -9.5973095
(switching optimization to BFGS)
Iteration 5:   log likelihood = -9.5850964
Iteration 6:   log likelihood = -9.5799049
Iteration 7:   log likelihood = -9.5770119
Iteration 8:   log likelihood = -9.5770099
Iteration 9:   log likelihood = -9.5770099

ARIMA regression

Sample:  1 - 4                                  Number of obs      =          4
                                                Wald chi2(1)       =     101.94
Log likelihood =  -9.57701                      Prob > chi2        =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .9759129    .096657    10.10   0.000     .7864686    1.165357
-------------+----------------------------------------------------------------
      /sigma |   1.812458   .8837346     2.05   0.020     .0803696    3.544545
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. arima y x,ar(1)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -4.3799561
Iteration 1:   log likelihood = -4.3799068  (backed up)
Iteration 2:   log likelihood =  -4.379678  (backed up)
Iteration 3:   log likelihood = -4.3796767  (backed up)
Iteration 4:   log likelihood = -4.3796761  (backed up)
(switching optimization to BFGS)
Iteration 5:   log likelihood = -4.3796757  (backed up)
Iteration 6:   log likelihood = -4.3235592
Iteration 7:   log likelihood = -4.2798453
Iteration 8:   log likelihood = -4.2471467
Iteration 9:   log likelihood =  -4.239353
Iteration 10:  log likelihood = -4.2384456
Iteration 11:  log likelihood =  -4.238435
Iteration 12:  log likelihood =  -4.238435

ARIMA regression

Sample:  1 - 4                                  Number of obs      =         4
                                                Wald chi2(2)       =   1001.98
Log likelihood = -4.238435                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
y            |
           x |    .635658   .0583723    10.89   0.000     .5212505    .7500656
       _cons |   .6512199          .        .       .            .           .
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.5631492   2.177484    -0.26   0.796    -4.830939    3.704641
-------------+----------------------------------------------------------------
      /sigma |   .6656358   .7509811     0.89   0.188            0    2.137532
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

7 Qualitative Dependent Variable (質的従属変数)

1. **Discrete Choice Model (離散選択モデル)**

2. **Limited Dependent Variable Model (制限従属変数モデル)**

3. **Count Data Model (計数データモデル)**

Usually, the regression model is given by:

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

where y_i is a continuous type of random variable within the interval from $-\infty$ to ∞.

When y_i is discrete or truncated, what happens?

7.1 Discrete Choice Model (離散選択モデル)

7.1.1 Binary Choice Model (二値選択モデル)

Example 1: Consider the regression model:

$$y_i^* = X_i\beta + u_i, \qquad u_i \sim (0, \sigma^2), \qquad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as 0 or 1, i.e.,

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0, \\ 0, & \text{if } y_i^* \leq 0. \end{cases}$$

Consider the probability that y_i takes 1, i.e.,

$$P(y_i = 1) = P(y_i^* > 0) = P(u_i > -X_i\beta) = P(u_i^* > -X_i\beta^*) = 1 - P(u_i^* \leq -X_i\beta^*)$$

$$= 1 - F(-X_i\beta^*) = F(X_i\beta^*), \quad \text{(if the dist. of } u_i^* \text{ is symmetric.)},$$