例 3： $X_1, X_2, \cdots, X_n$ 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，

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

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，

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

となる。

$\log l(\lambda)$ を最大にする λ を求める。

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

したがって，λ について解くと，λ の最尤推定値 $\hat{\lambda}$ は，

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

となる。

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

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

となる。

$\hat{\lambda}$ の分布を求めると，n が大きいとき，

$$\hat{\lambda} \sim N(\lambda, \sigma_\lambda^2) \qquad \text{ただし，} \quad \sigma_\lambda^2 = -\frac{1}{\sum_{i=1}^{n} \mathbf{E}\left(\frac{\mathbf{d}^2 \log f(X_i;\lambda)}{\mathbf{d}\lambda^2}\right)}$$

となる。

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

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

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

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

$\mathbf{E}(X_i)$ **は次のように計算される。**

$$\mathbf{E}(X_i) = \sum_{x=0}^{\infty} x f(x;\lambda) = \sum_{x=0}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \lambda\frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \sum_{x'=0}^{\infty} \lambda\frac{\lambda^{x'} e^{-\lambda}}{x'!} = \lambda$$

途中で， $x' = x - 1$ **と** $\sum_{x=0}^{\infty} f(x;\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = 1$ **を使っている。**

(*) 離散型確率変数 X について，その確率関数を $f(x)$ とし，X のとる範囲を x_1，x_2，$\cdots$ とするとき，$\displaystyle\sum_{i=1}^{\infty} f(x_i) = 1$ となる（確率の総和は 1）。

よって，

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

となり，

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

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。

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

したがって，λ について解くと，λ の最尤推定値 $\hat{\lambda}$ は，

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

となる。

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

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

となる。

$\hat{\lambda}$ の分布を求めると，n が大きいとき，

$$\hat{\lambda} \sim N(\lambda, \sigma_\lambda^2)$$ ただし，$$\sigma_\lambda^2 = -\frac{1}{\sum_{i=1}^{n} \mathbf{E}\left(\frac{\mathbf{d}^2 \log f(X_i;\lambda)}{\mathbf{d}\lambda^2}\right)}$$

となる。

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

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

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

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

したがって，

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

を得る。

10.2 最尤法を回帰分析へ応用

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 $u_i \sim N(0, \sigma^2)$ を仮定する。

Y_i の分布について，$\mathbf{E}(Y_i)$ と $\mathbf{V}(Y_i)$ を求める。

$$\mathbf{E}(Y_i) = \mathbf{E}(\alpha + \beta X_i + u_i) = \alpha + \beta X_i + \mathbf{E}(u_i) = \alpha + \beta X_i$$

$$\mathbf{V}(Y_i) = \mathbf{V}(\alpha + \beta X_i + u_i) = \mathbf{V}(u_i) = \sigma^2$$

よって， Y_i **は，平均** $\alpha + \beta X_i$**，分散** σ^2 **の正規分布となる。**

したがって， Y_i **の密度関数は，**

$$f(Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2\Big)$$

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，**

$$f(Y_1, Y_2, \cdots, Y_n; \theta) = \prod_{i=1}^{n} f(Y_i; \theta)$$

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

となる。

これは $\theta = \begin{pmatrix} \alpha \\ \beta \\ \sigma^2 \end{pmatrix}$ **の関数となっている。**

よって，尤度関数は，

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

となる。

対数尤度関数は，

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

となる。

$\log l(\theta)$ を最大にするために，

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

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

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

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，

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

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。

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix}$$，$$\theta = \begin{pmatrix} \alpha \\ \beta \\ \sigma^2 \end{pmatrix}$$とする。n が大きいとき，

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

となる。

ただし，$\Sigma_\theta = \Big(\sum_{i=1}^n \mathbf{E}\Big[\Big(\frac{\partial \log f(Y_i;\theta)}{\partial \theta}\Big)\Big(\frac{\partial \log f(Y_i;\theta)}{\partial \theta}\Big)'\Big]\Big)^{-1} = -\Big(\sum_{i=1}^n \mathbf{E}\Big[\frac{\partial^2 \log f(Y_i;\theta)}{\partial \theta \partial \theta'}\Big]\Big)^{-1}$ とする。

Y_i の密度関数 $f(Y_i;\theta)$ の対数は，

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

となる。

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

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

$$
= \begin{pmatrix}
-\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{Y_i - \alpha - \beta X_i}{\sigma^4} \\[2ex]
-\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_i(Y_i - \alpha - \beta X_i)}{\sigma^4} \\[2ex]
-\dfrac{Y_i - \alpha - \beta X_i}{\sigma^4} & -\dfrac{X_i(Y_i - \alpha - \beta X_i)}{\sigma^4} & \dfrac{1}{2\sigma^4} - \dfrac{(Y_i - \alpha - \beta X_i)^2}{\sigma^6}
\end{pmatrix}
$$

期待値をとると，

$$
\mathrm{E}\Big(\frac{\partial^2 \log f(Y_i;\theta)}{\partial\theta\partial\theta'}\Big) = \mathrm{E}\begin{pmatrix}
-\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{Y_i-\alpha-\beta X_i}{\sigma^4} \\[3ex]
-\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_i(Y_i-\alpha-\beta X_i)}{\sigma^4} \\[3ex]
-\dfrac{Y_i-\alpha-\beta X_i}{\sigma^4} & -\dfrac{X_i(Y_i-\alpha-\beta X_i)}{\sigma^4} & \dfrac{1}{2\sigma^4}-\dfrac{(Y_i-\alpha-\beta X_i)^2}{\sigma^6}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
-\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & 0 \\[3ex]
-\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & 0 \\[3ex]
0 & 0 & -\dfrac{1}{2\sigma^4}
\end{pmatrix}
$$

となる。

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

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

(*) **正方行列** A**，**B **について，** $\begin{pmatrix} A & 0 \\ 0 & B\end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1}\end{pmatrix}$ **となる。**

したがって，

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

となる。

$\longrightarrow$ 最小二乗推定量の分布と同じ。

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

$$
= \sigma^2 \begin{pmatrix} \dfrac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} & -\dfrac{\overline{X}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \\ -\dfrac{\overline{X}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} & \dfrac{1}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \end{pmatrix}
$$

$$
= \begin{pmatrix} \dfrac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \overline{X})^2} & -\dfrac{\sigma^2\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ -\dfrac{\sigma^2\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} & \dfrac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \end{pmatrix}
$$

すなわち，

$$
\hat{\alpha} \ \sim \ N(\alpha, \ \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \overline{X})^2}) \qquad \hat{\beta} \ \sim \ N(\beta, \ \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2})
$$

となる。