

2014年後期 応用確率統計

⑫ 確率過程

河野 行雄

kawano@pe.titech.ac.jp

2015年1月22日

復習

■ 仮説

帰無仮説 H_0 : 「 $\theta_1 = \theta_2$ 」 「 $\theta = \theta_0$ 」

対立仮説 H_1 : 「 $\theta_1 \neq \theta_2$ 」 「 $\theta = \theta_1$ 」

■ 検定統計量

$$T \equiv \sqrt{\frac{nm}{n+m}} \frac{\bar{Y} - \bar{X}}{\sigma} \quad \longleftarrow \text{分散1の正規分布}$$

■ 有意水準

$$P(|T| \geq \mu) = \alpha \quad \longleftarrow \text{第一種の過誤}$$

■ 仮説の受容と棄却

$|T| < \mu \quad \longrightarrow \quad$ 帰無仮説を受容 「 $\theta_1 = \theta_2$ 」

$|T| \geq \mu \quad \longrightarrow \quad$ 帰無仮説を棄却 「 $\theta_1 \neq \theta_2$ 」

確率過程とは？

■ 多変量統計学(解析)

観測データ: $Y_i(X) = \theta(X) + \varepsilon_i \quad (i = 1, \dots, n)$

ε_i は i.i.d. で正規分布 $N(0, \sigma^2)$ に従う

未知母数は X の関数

2つの確率変数(変量) X と Y の関係を統計的に解析

■ 確率過程

X が時間 t のとき $Y_i(t)$ を確率過程と呼ぶ

例: 株価、会話、雑音、交通量

回帰直線

■ 回帰曲線

2つの変量 X と Y の関係を表す曲線

$$y = E(Y(x)) = \theta(x)$$

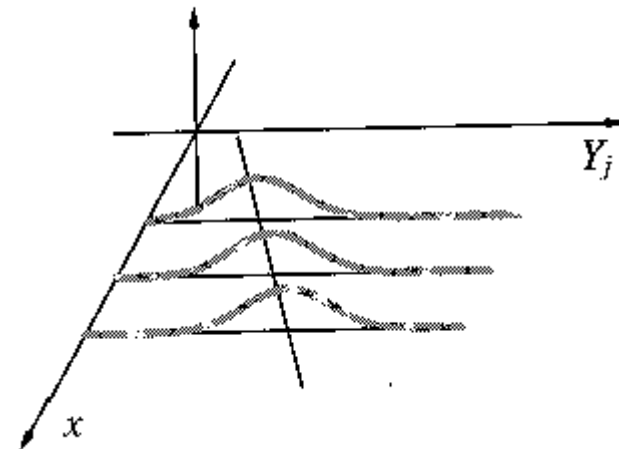
■ 回帰直線

$$\theta(x) = ax + b$$

回帰係数
↙ ↘
↖ ↗

$x = t$ のとき: 回帰 $\doteq$ 予測

例: 温度と抵抗率、緯度と気温、努力と成績



尤度関数

■ 正規分布

$$f(Y_i, a, b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - aX_i - b)^2}{2\sigma^2}\right)$$

■ 尤度関数

$$\begin{aligned} L(a, b) &= \prod_{i=1}^n f(Y_i, a, b) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - aX_i - b)^2\right) \end{aligned}$$

$$\log L(a, b) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - aX_i - b)^2$$

二乗誤差

対数尤度 $\log L(a, b)$ の最大化 = 二乗誤差の最小化

回帰係数の最尤推定

■ 最尤推定

$$(\hat{a}, \hat{b}) = \arg \max_{a, b} \log L(a, b)$$

$$\frac{\partial}{\partial a} \log L(a, b) = -\sum_{i=1}^n X_i Y_i + b \sum_{i=1}^n X_i + a \sum_{i=1}^n X_i^2 = 0$$

$$\frac{\partial}{\partial b} \log L(a, b) = -\sum_{i=1}^n Y_i + nb + a \sum_{i=1}^n X_i = 0$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \Downarrow \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$\hat{a} = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2} \quad \hat{b} = \bar{Y} - \hat{a} \bar{X}$$

不偏分散と不偏共分散

■ 回帰係数の最尤推定量

$$\begin{aligned}\hat{a} &= \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\hat{S}_{XY}}{\hat{S}_X^2}\end{aligned}$$

■ 不偏分散と不偏共分散

不偏分散

$$\hat{S}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

不偏共分散

$$\hat{S}_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

共分散と相関

■ 共分散

2つの変量 X と Y の相関関係を調べる指標

$$\hat{S}_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

■ 相関係数

共分散を X と Y の標準偏差で正規化 $\longrightarrow |\hat{\rho}| \leq 1$

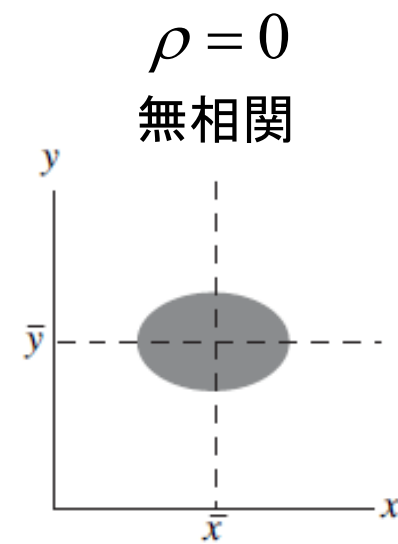
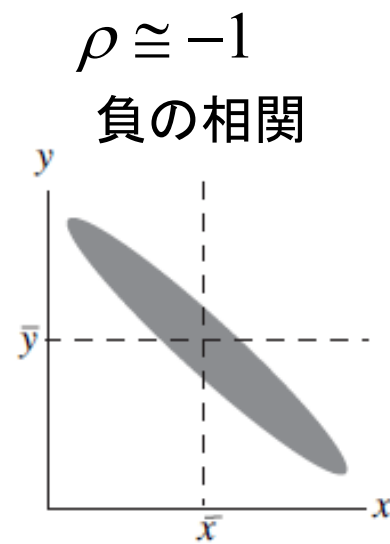
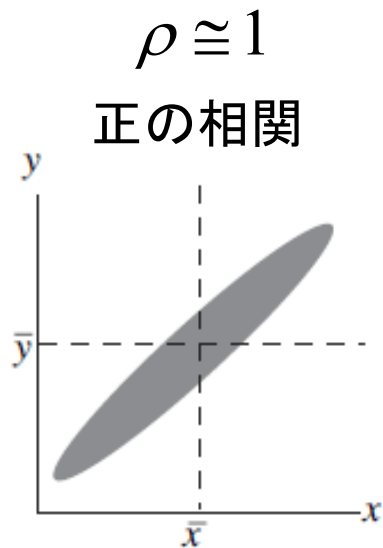
$$\hat{\rho} = \frac{\hat{S}_{XY}}{\hat{S}_X \hat{S}_Y} \xrightarrow{n \rightarrow \infty} \rho = \frac{E((X - m_X)(Y - m_Y))}{\sqrt{V(X)}\sqrt{V(Y)}}$$

相関係数

■ 相関係数

$$\rho = \frac{E((X - m_X)(Y - m_Y))}{\sqrt{V(X)}\sqrt{V(Y)}} \quad -1 \leq \rho \leq 1$$

■ 散布図と相関係数



自己相関

■ 自己共分散

確率過程 $Y(t) = \theta(t) + \varepsilon$ に対して

$$S_{t_1 t_2} = E(((Y(t_1) - \theta(t_1))(Y(t_2) - \theta(t_2))))$$

■ 自己相関

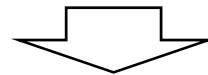
$$\rho_{t_1 t_2} = \frac{S_{t_1 t_2}}{S_{t_1} S_{t_2}} = \frac{S_{t_1 t_2}}{\sigma^2} \quad -1 \leq \rho_{t_1 t_2} \leq 1$$

定常過程

■ 定常過程

平均値が時間によらず一定: $E(Y(t)) = \theta(t) = \theta$

共分散が時間差のみの関数: $\rho_{t_1 t_2} = \frac{S_{t_1 t_2}}{\sigma^2} = \frac{S_{\Delta t}}{\sigma^2} \quad \Delta t = t_2 - t_1$



$Y(t)$ は定常過程 $\longrightarrow$ 確率(集合)平均と時間平均が一致

例: サイコロ $\rightarrow$ 細工をしない限り定常

ルーレット $\rightarrow$ ディーラーが交代しない限り定常

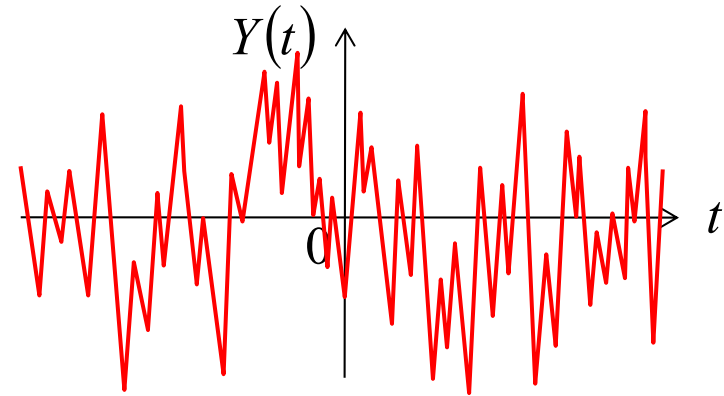
雑音 $\rightarrow$ 温度が変化しない限り定常

雜音

■ 平均、分散

$$E(Y(t)) = 0$$

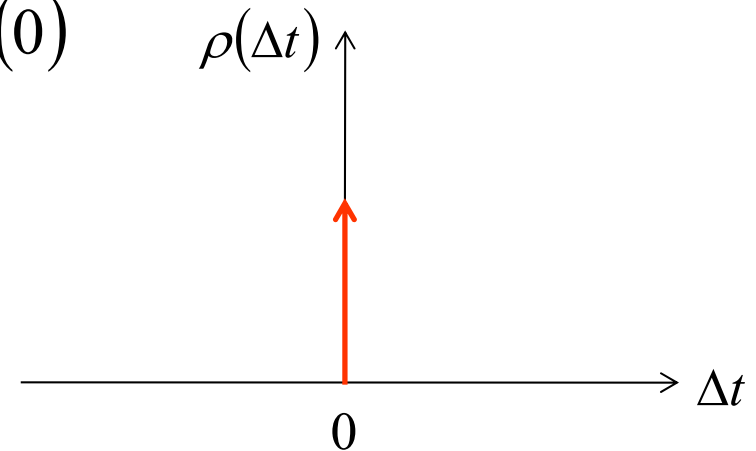
$$V(Y(t)) = E(Y^2(t)) = \sigma^2$$



■ 共分散、自己相關

$$S(\Delta t) = E(Y(t)Y(t + \Delta t)) = \sigma^2 \delta(0)$$

$$\rho(\Delta t) = \frac{S(\Delta t)}{V(Y(t))} = \delta(0)$$



ランダムパルス

- ランダムパルス 矩形パルス関数

$$Y(t) = \sum_{i=-\infty}^{\infty} m_i \text{rect}(t - iT)$$

$$m_i : +A \text{ or } -A$$

- 平均、分散

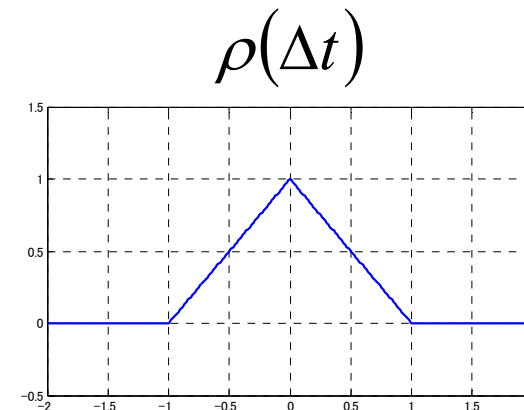
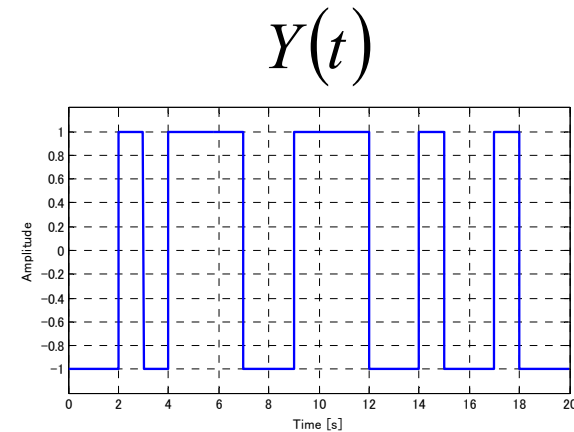
$$E(Y(t)) = 0$$

$$V(Y(t)) = E(Y^2(t)) = A^2$$

- 共分散、自己相関

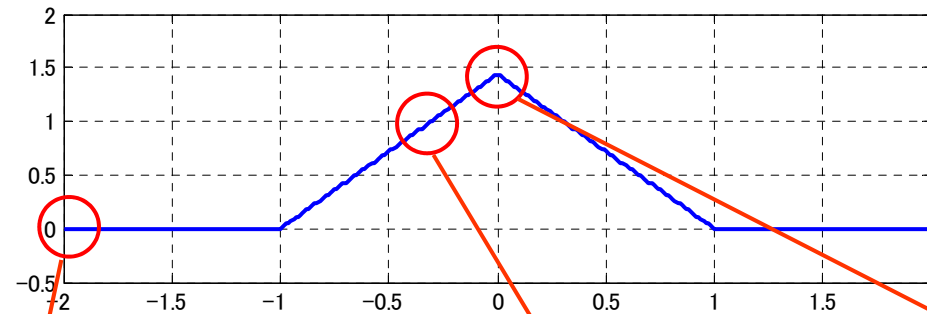
$$S(\Delta t) = \begin{cases} A^2(1 - |\Delta t|/T) & |\Delta t| < T \\ 0 & |\Delta t| \geq T \end{cases}$$

$$\rho(\Delta t) = \frac{S(\Delta t)}{V(Y(t))}$$



ランダムパルス信号(自己相関)

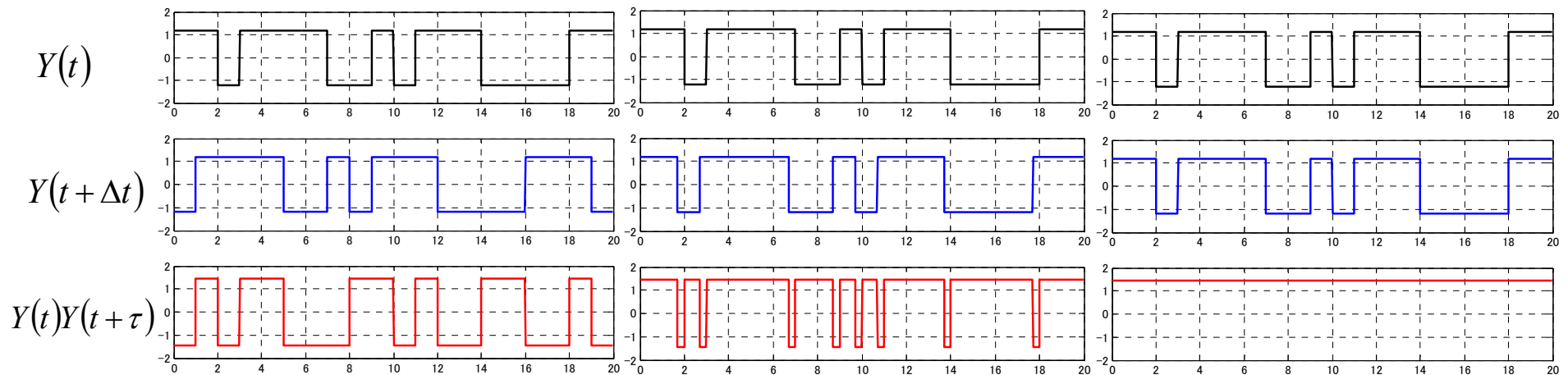
$$S(\Delta t) = E[Y(t)Y(t + \Delta t)]$$



$$|\Delta t| = 2T$$

$$|\Delta t| = 0.3T$$

$$|\Delta t| = 0$$



まとめ

■ 多変量統計学

$$Y_i(X) = \theta(X) + \varepsilon_i$$

$X = t$ のとき $\longrightarrow$ 確率過程

■ 回帰直線

$$\theta(x) = ax + b$$

$$\hat{a} = \frac{\hat{S}_{XY}}{\hat{S}_X^2} \quad \hat{b} = \bar{Y} - \hat{a}\bar{X}$$

■ 共分散と相関係数

$$\hat{S}_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad \hat{S}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\hat{\rho} = \frac{\hat{S}_{XY}}{\hat{S}_X \hat{S}_Y}$$