科学研究費シンポジウム
日本女子大学、2009年10月2日（金）

# Testing quasi-independence for truncation data

江村剛志, Weijing Wang,
国立交通大学、統計学研究所

# 発表内容

1. 切断データ（Review)
2. Quasi-independence（Review)
3. 提案する方法
4. 統計量の漸近分布
5. 今後の課題

# 切断データ（Review）

**例1**；Channing House data
(Hyde, 1980, *Biostatistics Casebook*, pp.31-46)
高齢者福祉施設で入局者の生存時間を長期的に観測

$$\begin{cases} X & \text{施設入局時の年齢} \\ Y & \text{死亡時の年齢} \end{cases}$$

$$* Y - X : \text{Residual life time}$$

- Hydeは $X \perp Y$ の下で $Y$ の分布を推定
- 観測値が得られる条件；$X \leq Y$
- $X$ は $Y$ を左から切断(Left-truncation)

# 切断データ（Review）

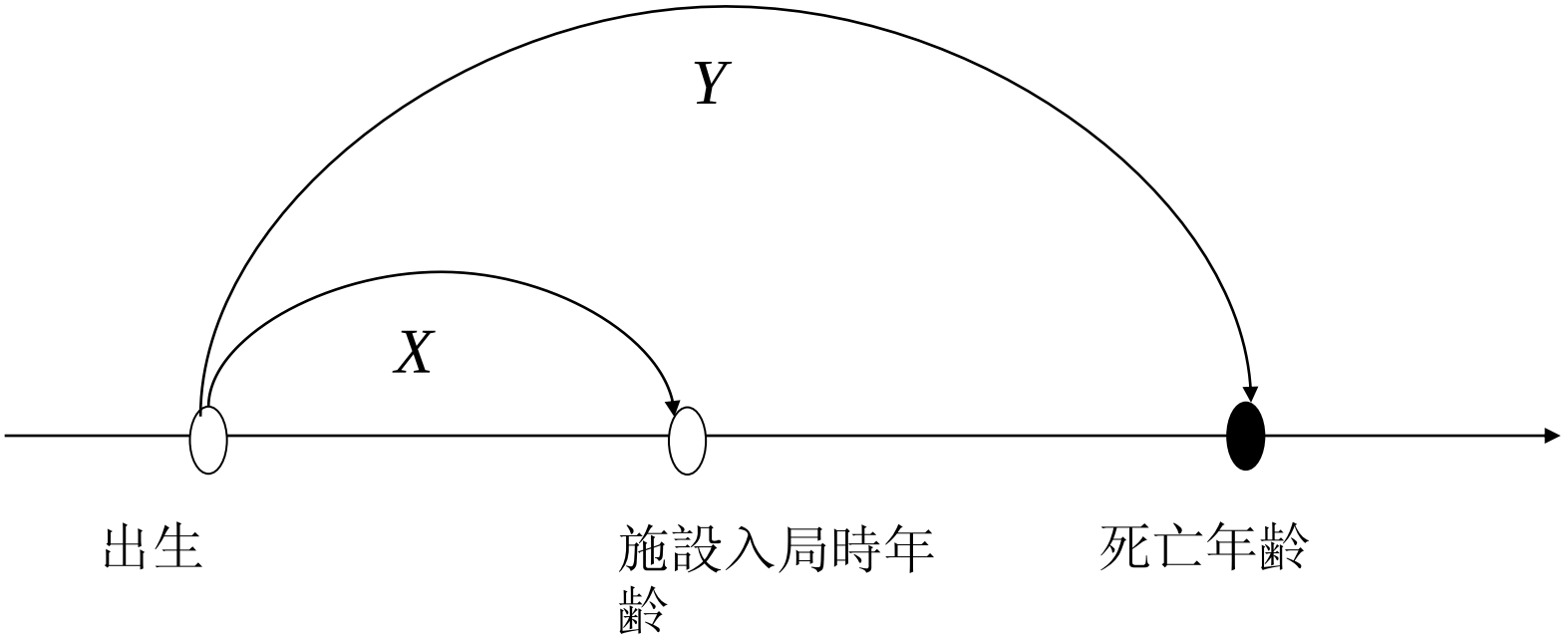



図１、Channing House data (Hyde, 1980)

観測値 $\{(X_j, Y_j)\,(j = 1, \ldots, n)\}$ subject to $X_j \leq Y_j$

# 切断データ（Review）

**例2**；AIDS data (Kalbfleisch & Lawless, 1989, JASA)

AIDSの潜伏期間を一定期間102ヶ月に渡って調査

$$\begin{cases} T & \text{ある時点から感染までの時間(月)} \\ X & \text{感染から発症までの時間(月)} \end{cases}$$

$$*\text{観測可能条件}: X \le T$$

- 観測値が得られる条件；$X \le Y$

$$\text{ここで、} Y = 102 - T$$

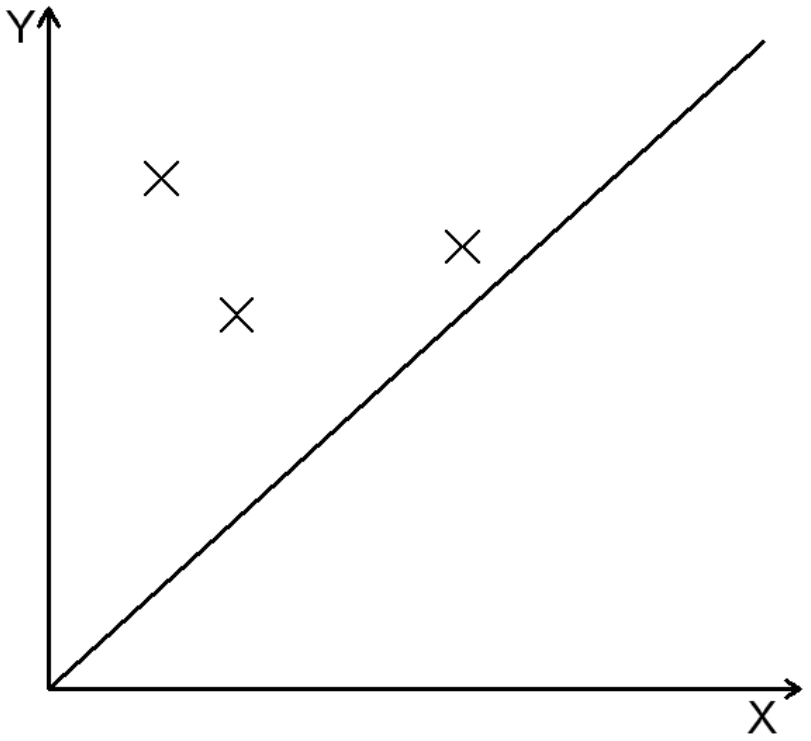
- $Y$ は $X$ を右から切断(Right-truncation)
- Lagakos et al. (1988)は $X \perp Y$ の下で $X$ の分布関数の推定量を提案

# 切断データ（Review）

- 切断データ　$\{(X_j, Y_j)\,(j = 1, \ldots, n)\}$
  Subject to　$X_j \le Y_j$

- Left-truncation, Right-truncation共に上記のデータ構造

- $X \perp Y$が仮定される (Kalbfleish & Lawless; Hyde)



- 独立性の仮説は実データで成立するか？

  In general, no

- 独立性仮説は統計的に検定可能か？

  Yes, but quasi-independence

# Quasi-independence (Review)

- 完全データ $(X_1, Y_1), \ldots, (X_n, Y_n) \sim F(x, y)$

$$\tilde{H}_0 : F(x, y) = F_X(x) F_Y(y)$$

1) FisherのZ検定（ノンパラメトリック）

$$\rho = \frac{\sum_j (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_j (X_j - \bar{X})^2 \sum_j (Y_j - \bar{Y})^2} \qquad \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right) \sim N(0,1)$$

2) 尤度比検定（パラメトリック、漸近有効）

$$2 \log L / L_0 \sim \chi^2_{df=1}$$

# Quasi-independence（Review）

切断データにおける独立性をどのように定式化するか？

- データ　$(X_1, Y_1), \ldots, (X_n, Y_n)$

  $\sim F_c(x, y) = \Pr(X \le x, Y \le y \mid X \le Y)$

- 独立性仮説　$\tilde{H}_0 : \Pr(X \le x, Y \le y) = F_X(x) F_Y(y)$

  は切断データから検定不可能

- 定義；Quasi-independence

$$H_0 : \Pr(X \le x, Y > y \mid X \le Y) = F_X(x) S_Y(y) / c_0$$

ここで，$c_0 = -\iint_{u \le v} dF_X(u) dS_Y(v)$

# Quasi-independence（Review）

- 目的;切断データ

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

$$\sim F_c(x, y) = \Pr(X \le x, Y \le y \mid X \le Y)$$

よりQuasi-independence

$$H_0 : \Pr(X \le x, Y > y \mid X \le Y) = F_X(x) S_Y(y) / c_0$$

の統計的検定を行う

- 簡潔に述べると、Quasi-independenceとは条件付分布関数が2つの関数の積で書けること
- Reject　Quasi-independence

　→　Reject Independence

（完全データ;Reject　ρ＝0　→　Reject　Independence）

# 提案する方法

- Emura and Wang (2009), J of multivariate analysis より紹介
- Notation; 計数過程

$$\Delta(x,y)=\sum_{j=1}^{n} I(X_j=x, Y_j=y),$$

$$N_{\bullet 1}(x,dy)=\sum_{i=1}^{n} I(X_i\le x, Y_i=y)$$

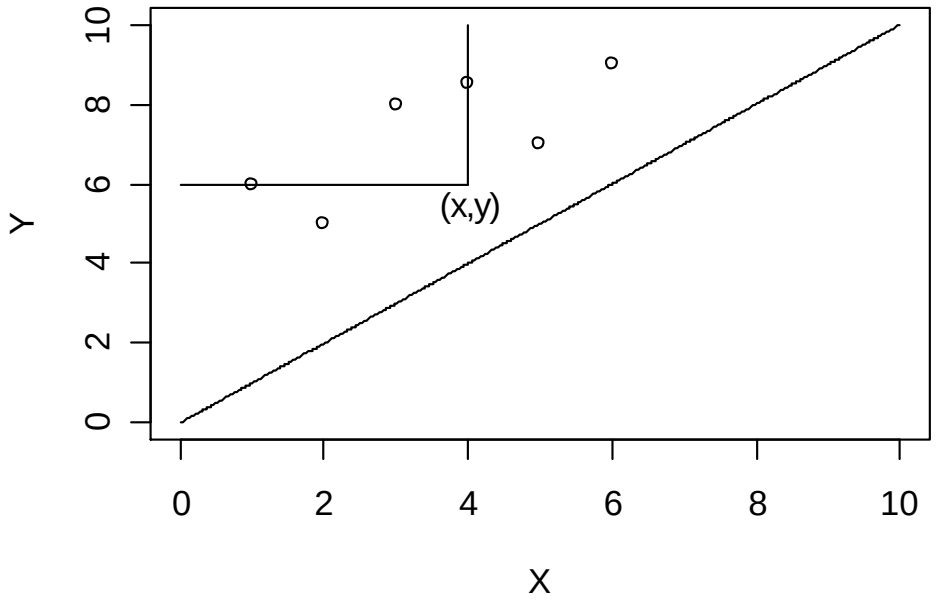
$$N_{1\bullet}(x,dy)=\sum_{i=1}^{n} I(X_i=x, Y_i\ge y)$$

$$R(x,y)=\sum_{j=1}^{n} I(X_j\le x, Y_j\ge y)$$

- ２×２集計表

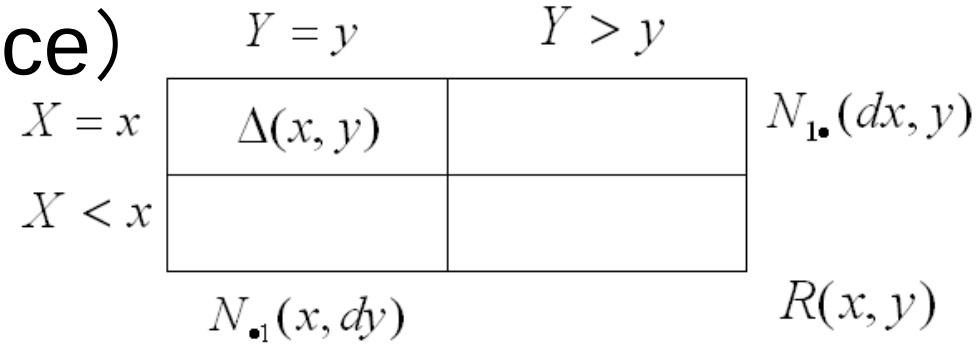
| | $Y=y$ | $Y>y$ | |
|---|---|---|---|
| $X=x$ | $\Delta(x,y)$ | | $N_{1\bullet}(dx,y)$ |
| $X<x$ | | | |
| | $N_{\bullet 1}(x,dy)$ | | $R(x,y)$ |

**Truncation Data**



# 提案する方法

- 帰無仮説 $H_0$
  （Quasi-independence）
  の下で $\Delta(x,y)$
  は超幾何分布に従う




$$E\{\Delta(x,y)\,|\,\text{marginal counts}\} = \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}$$

- ログランク型統計量

$$L_w = \iint_{x\le y} W(x,y)\left[\Delta(x,y) - \frac{N_{1\bullet}(dx,y)N_{\bullet 1}(x,dy)}{R(x,y)}\right]$$

How to choose W(x,y)？

# 統計量の漸近分布

- $G^{\rho}$ クラス統計量（Fleming-Harington型）

$$L_{\rho} = \iint_{x \le y} \hat{\pi}(x, y-)^{\rho} \left\{ \Delta(dx, dy) - \frac{N_{1\bullet}(dx, y) N_{\bullet 1}(x, dy)}{R(x, y)} \right\}$$

$$\rho \ge 0; \text{定数} \qquad \hat{\pi}(x, y-) = \frac{R(x, y)}{n}; \text{ Number at risk}$$

- 漸近正規性（参考；Emura and Wang（2009））

$$L_{\rho} \underset{H_0}{\to} N(0, \sigma_{\rho}^2)$$

- Plug-in分散推定量

$$\sigma_{\rho}^2 \approx \sum_j \left[ \frac{1}{n} \sum_k I\{A_{jk}\} \hat{\pi}(\tilde{X}_{jk}, \tilde{Y}_{jk}-)^{\rho-1} \operatorname{sgn}\{(X_j - X_k)(Y_j - Y_k)\} + \frac{(\rho+1) L_{\rho}}{n} \right.$$

$$\left. + \frac{\rho-1}{n^2} \sum_{k<l} I\{A_{kl}\} \hat{\pi}(\tilde{X}_{kl}, \tilde{Y}_{kl}-)^{\rho-2} \operatorname{sgn}\{(X_k - X_l)(Y_k - Y_l)\} I(X_j \le \tilde{X}_{kl}, Y_j \ge \tilde{Y}_{kl}) \right]^2.$$

# 統計量の漸近分布

★分散推定量はJackknifeの方が優れている（千葉大学テクニカルレポートVol.24、No.12、2008）

- 計算がプラグイン推定量よりも容易である
- $G^{\rho}$クラス以外にも適用可能
- 漸近近似の精度がプラグイン推定量よりも良い
- 分散の推定量の一致性

（一致性の証明;汎関数表示の連続ガトー可微分性を用いる）

結論；$|L_{\rho}/\hat{\sigma}_{\rho}^{2}|>1.96$ ならば $H_0$ を棄却

$$\hat{\sigma}_{\rho}^{2}=n/(n-1)\sum_{j}(L_{\rho}^{(-j)}-L_{\rho}^{(\cdot)})^{2}$$

# 統計量の漸近分布

- 統計量のPower評価基準；Locally most powerfuII
  対立仮説が帰無仮説に近づくときの最適性
- 対立仮説の列（ $H_n \to H_0$ Local alternative）

$$H_n : \Pr(X \le x, Y > y \mid X \le Y) = \phi_{\alpha_n}^{-1}[\{\phi_{\alpha_n}\{F_X(x)\} + \phi_{\alpha_n}\{S_Y(y)\}]/c$$

$$\alpha_n = 1 + 1/\sqrt{n}$$

$$1.\ \phi_\alpha(t) = (t^{-(\alpha-1)} - 1)/(\alpha - 1); \text{Clayton - Copula model}$$

$$2.\ \phi_\alpha(t) = \log\{(1-\alpha)/(1-\alpha^t)\}; \text{Frank - Copula model}$$

# 統計量の漸近分布

- $H_n \to H_0$ のとき、 $L_\rho \underset{H_n}{\to} N(\mu_\rho, \sigma_\rho^2)$

  ただし、$\mu_\rho$は解析的に書けない（Emura and Wang, 2009）

- Powerを最大化する

  $\leftrightarrow \quad \tilde{\mu}_\rho^2 = \mu_\rho^2 / \sigma_\rho^2$ を最大化する $\rho$ を選択する

$$1.\ \phi_\alpha(t) = (t^{-(\alpha-1)} - 1)/(\alpha - 1); \text{Clayton-Copula model}$$

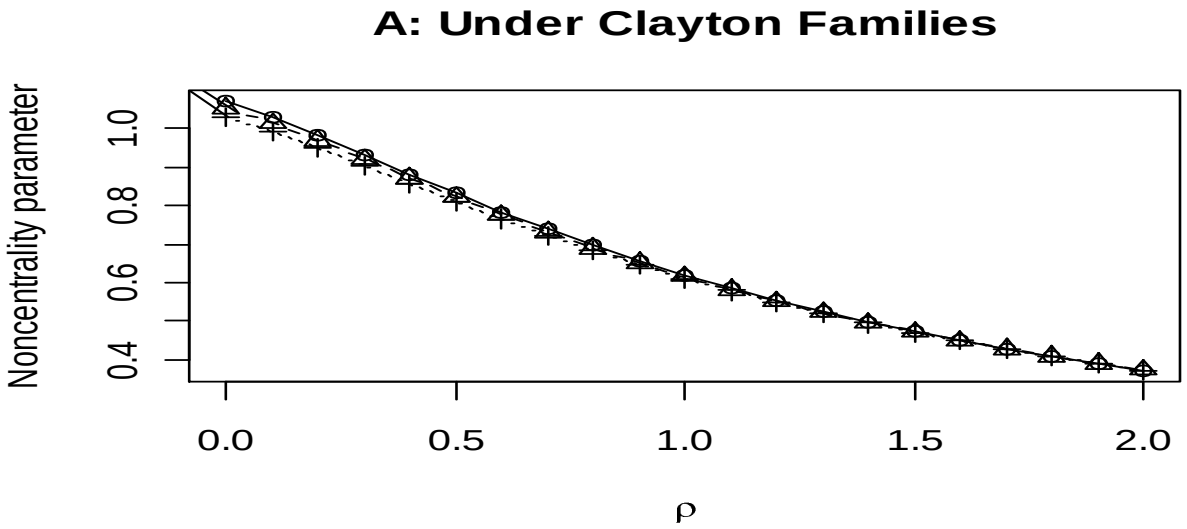
$$\Longrightarrow \rho = 0 \text{ で } \tilde{\mu}_\rho \text{ が最大化}$$

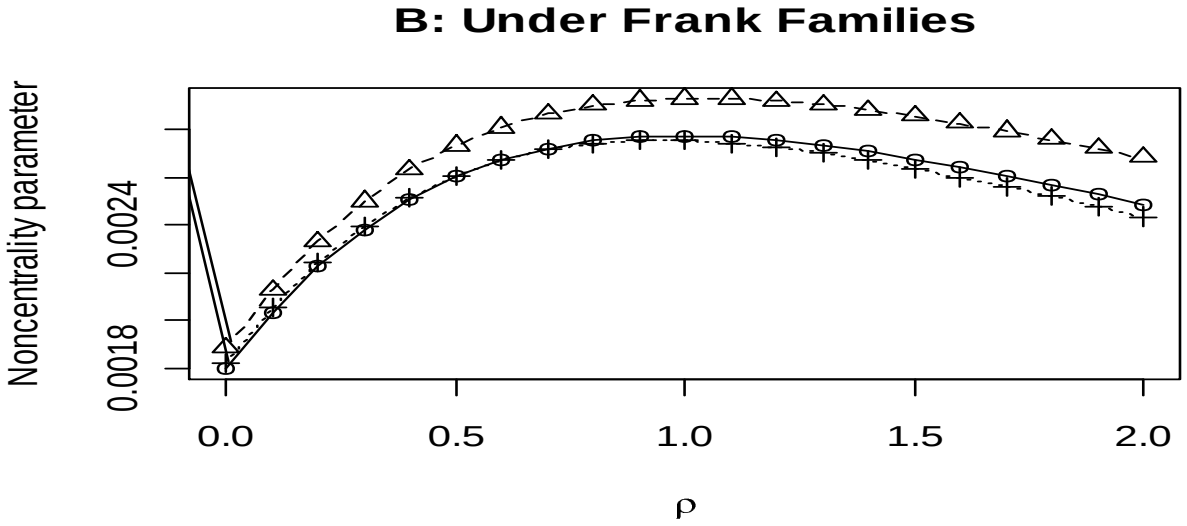
$$2.\ \phi_\alpha(t) = \log\{(1-\alpha)/(1-\alpha^t)\}; \text{Frank-Copula model}$$

$$\Longrightarrow \rho = 1 \text{ で } \tilde{\mu}_\rho \text{ が最大化}$$

ただし、$\mu_\rho$は解析的に書けないのでMonteCarlo近似した

# 統計量の漸近分布


A: Under Clayton Families
Noncentrality parameter
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
ρ


B: Under Frank Families
Noncentrality parameter
0.0018
0.0024
0.0
0.5
1.0
1.5
2.0
ρ

# 今後の課題

- 両側切断データへの拡張

  Xが右切断と左切断を両方受ける場合

  両側切断データ;

  $$(X_1, L_1, R_1), \ldots, (X_n, L_n, R_n)$$

  $$\sim H(x, l, r) = \Pr(X \le x, L \le l, R \le r \mid L \le X \le R)$$

  *仮説；* $H_0 : X \perp (L, R)$

- 先行研究1;Efron and Petrosian（1994，1999;JASA）
  ★宇宙上の星の明るさの測定値が両側切断される
- 先行研究2;Martin and Betensky (2005)

  ★両側切断データの下で、Quasi-independenceをKendallのTauを用いて検定する手法を提案

# ご静聴ありがとうございました