Marginal quantiles:
Asymptotics for functions of order statistics.

G. Jogesh Babu

Penn State University

Outline

Data depth

- Data depth provides an ordering of all points from the center outward. Contours of depth are often used to reveal the shape and structure of a multivariate data set. The depth of a point x in a one-dimensional data set $\{x_1, x_2, \cdots, x_n\}$ can be defined as the minimum of the number of data points on one side of x. [J. W. Tukey, in Proceedings of the International Congress of Mathematicians (1974), Vol. 2].

Data depth

- Data depth provides an ordering of all points from the center outward. Contours of depth are often used to reveal the shape and structure of a multivariate data set. The depth of a point x in a one-dimensional data set $\{x_1, x_2, \cdots, x_n\}$ can be defined as the minimum of the number of data points on one side of x. [J. W. Tukey, in Proceedings of the International Congress of Mathematicians (1974), Vol. 2].

- Several multidimensional depth measures $D_n(x; x_1, \cdots, x_n)$ for $x \in R^k$ were considered by many that satisfy certain mathematical conditions. If the data is from a spherical or elliptic distribution, the depth contours are generally required to converge to spherical or elliptic shapes.

Marginal quantiles

Tests of significance for population medians based on marginal quantiles can be derived from the joint distribution of marginal sample quantiles.

Marginal quantiles

Tests of significance for population medians based on marginal quantiles can be derived from the joint distribution of marginal sample quantiles.

F a k-dimensional distribution function

F_j denotes the j-th marginal distribution function.

$F_j^{-1}(u) = \inf\{x : F_j(x) \geq u\}$, $0 < u < 1$
u-th quantile of the jth marginal.

Theorem

$\mathbf{X}_1, \ldots, \mathbf{X}_n \;\; i.i.d. \;\sim$ **F**, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$.
$X_{ij} \sim F_j$
$0 < q_1, \ldots, q_k < 1$. $\qquad \delta_j$ density of F_j at $F_j^{-1}(q_j)$.

$\hat{\theta}_j$ denotes the q_j-th sample quantile based on $X_{1j}, \ldots, X_{nj}$

Theorem

$\mathbf{X}_1, \ldots, \mathbf{X}_n \;\; i.i.d. \; \sim$ **F**, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$.
$X_{ij} \sim F_j$
$0 < q_1, \ldots, q_k < 1$. $\quad \delta_j$ density of F_j at $F_j^{-1}(q_j)$.

$\hat{\theta}_j$ denotes the q_j-th sample quantile based on $X_{1j}, \ldots, X_{nj}$

Theorem

Let F_j be twice continuously differentiable in a neighborhood of $F_j^{-1}(q_j)$ and $\delta_j > 0$. Then

$$\sqrt{n}(\hat{\theta}_1 - F_1^{-1}(q_1), \ldots, \hat{\theta}_k - F_k^{-1}(q_k)) \sim N(\mathbf{0}, \Sigma)$$

Covariance Matrix

The covariance matrix Σ is given by

$$\begin{pmatrix} \frac{q_1(1-q_1)}{\delta_1^2} & \sigma_{12} & \cdots & \sigma_{1k} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \frac{q_k(1-q_k)}{\delta_k^2} \end{pmatrix},$$

where for $i \neq j$, $\sigma_{ij} = (F_{ij}(F_i^{-1}(q_i), F_j^{-1}(q_j)) - q_i q_j)/(\delta_i \delta_j)$

Covariance Matrix

The covariance matrix Σ is given by

$$\begin{pmatrix} \frac{q_1(1-q_1)}{\delta_1^2} & \sigma_{12} & \cdots & \sigma_{1k} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \frac{q_k(1-q_k)}{\delta_k^2} \end{pmatrix},$$

where for $i \neq j$, $\sigma_{ij} = (F_{ij}(F_i^{-1}(q_i), F_j^{-1}(q_j)) - q_i q_j)/(\delta_i \delta_j)$

In practice σ_{ij} can be estimated by bootstrap $E^*(n(\theta_i^* - \hat{\theta}_i)(\theta_j^* - \hat{\theta}_j))$, where θ_j^* denotes the bootstrapped marginal sample quantile.

Quantile process

For $(q_1, \ldots, q_k) \in (0,1)^k$, define the sample quantile process,
$Z_n(q_1, \ldots, q_k) = \sqrt{n}[\delta_1(\hat{\theta}_1 - F_1^{-1}(q_1)), \ldots, \delta_k(\hat{\theta}_k - F_k^{-1}(q_k))]$

Quantile process

For $(q_1, \ldots, q_k) \in (0,1)^k$, define the sample quantile process,
$Z_n(q_1, \ldots, q_k) = \sqrt{n}[\delta_1(\hat{\theta}_1 - F_1^{-1}(q_1)), \ldots, \delta_k(\hat{\theta}_k - F_k^{-1}(q_k))]$

Theorem

Suppose for $j = 1, \ldots, k$, the marginal d.f. F_j is twice differentiable on (a_j, b_j), where

$$-\infty \leq a_j = \sup\{x : F_j(x) = 0\}$$

$$\infty \geq b_j = \inf\{x : F_j(x) = 1\}$$

If the first two derivatives of F_j satisfy some regularity conditions, then $Z_n(q_1, \ldots, q_k)$ converges weekly to a Gaussian random element $(W_1, \ldots, W_k)$ on $C[0,1]^k$

$f_j \neq 0$ on (a_j, b_j),

$$\max_{1 \leq j \leq k} \sup_{a_j < x < b_j} F_j(x)(1 - F_j(x)) \frac{|f_j'(x)|}{f_j^2(x)} < \infty$$

Each marginal of Z_n converges weakly to a Brownian bridge. The covariance is given by

$$E(W_i(t)W_j(s)) = P(F_i(X_{i1}) \leq t, F_j(X_{j1}) \leq s) - ts$$

$f_j \neq 0$ on (a_j, b_j),

$$\max_{1 \leq j \leq k} \sup_{a_j < x < b_j} F_j(x)(1 - F_j(x)) \frac{|f_j'(x)|}{f_j^2(x)} < \infty$$

Each marginal of Z_n converges weakly to a Brownian bridge.
The covariance is given by

$$E(W_i(t)W_j(s)) = P(F_i(X_{i1}) \leq t, F_j(X_{j1}) \leq s) - ts$$

– *Collaboration with C. R. Rao*

Regression

How to estimate the regression coefficients when the association among the paired data is partially or completely lost?

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

X_i i.i.d with mean μ and standard deviation σ_X, ϵ_i i.i.d with mean zero and standard deviation σ_ϵ and $\{X_i\}$ and $\{\epsilon_i\}$ are independent sequences. P_n the set of all permutations of $\{1, \ldots, n\}$.

$$\Phi(a, b) = \min_{\pi \in P_n} \sum_{i=1}^{n} (Y_{\pi(i)} - a - bX_i)^2$$

$$\sum (Y_i - a - bX_i)^2 \geq S_{YY}(1 - r_{XY}^2).$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{a,b} \Phi(a, b)$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{\pi \in P_n} S_{YY}(1 - r^2_{X\pi(Y)})$$

The permutation that minimizes Φ is free from α, β.

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{a,b} \Phi(a, b)$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{\pi \in P_n} S_{YY}(1 - r^2_{X\pi(Y)})$$

The permutation that minimizes Φ is free from α, β.

The main difficulty is the computational complexity.
$n!$ permutations requiring that many computations.

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{a,b} \Phi(a, b)$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{\pi \in P_n} S_{YY}(1 - r^2_{X\pi(Y)})$$

The permutation that minimizes Φ is free from α, β.

The main difficulty is the computational complexity.
$n!$ permutations requiring that many computations.

$\Phi(\hat{\alpha}, \hat{\beta})$ depends on permutation π through $\max_{\pi \in P_n} r^2_{X\pi(Y)}$.

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{a,b} \Phi(a, b)$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{\pi \in P_n} S_{YY}(1 - r^2_{X\pi(Y)})$$

The permutation that minimizes Φ is free from α, β.

The main difficulty is the computational complexity.
$n!$ permutations requiring that many computations.

$\Phi(\hat{\alpha}, \hat{\beta})$ depends on permutation π through $\max_{\pi \in P_n} r^2_{X\pi(Y)}$.

How to identify the permutation that maximizes the value?

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{a,b} \Phi(a, b)$$

$$\Phi(\hat{\alpha}, \hat{\beta}) = \min_{\pi \in P_n} S_{YY}(1 - r^2_{X\pi(Y)})$$

The permutation that minimizes Φ is free from α, β.

The main difficulty is the computational complexity.
$n!$ permutations requiring that many computations.

$\Phi(\hat{\alpha}, \hat{\beta})$ depends on permutation π through $\max_{\pi \in P_n} r^2_{X\pi(Y)}$.

How to identify the permutation that maximizes the value?

$R_X(i)$ is the rank of X_i in $\{X_1, \ldots, X_n\}$

$\bar{R}_X(i) = n - R_X(i) + 1$. Define R_Y and $\bar{R}_Y$ similarly.

$\pi_1 = R_Y^{-1} R_X$ and $\pi_2 = \bar{R}_Y^{-1} R_X$

Lemma

$$\Phi(\hat{\alpha}, \hat{\beta}) = S_{YY}(1 - \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\}).$$

$$\hat{\beta}^2 = \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\}S_{YY}/S_{XX}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

$\frac{1}{n}\sum_{i=1}^n X_{(i)}Y_{(i)}$ and $\frac{1}{n}\sum_{i=1}^n X_{(i)}Y_{(n-i+1)}$ are used in computing $\hat{\beta}$.

Lemma

$$\Phi(\hat{\alpha}, \hat{\beta}) = S_{YY}(1 - \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\}).$$

$$\hat{\beta}^2 = \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\} S_{YY}/S_{XX}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(i)}$ and $\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(n-i+1)}$ are used in computing $\hat{\beta}$. Their limits lead to that of $\hat{\beta}$, and would aid in the estimation of the bias of $\hat{\beta}$.

Lemma

$$\Phi(\hat{\alpha}, \hat{\beta}) = S_{YY}(1 - \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\}).$$

$$\hat{\beta}^2 = \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\} S_{YY}/S_{XX}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(i)}$ and $\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(n-i+1)}$ are used in computing $\hat{\beta}$. Their limits lead to that of $\hat{\beta}$, and would aid in the estimation of the bias of $\hat{\beta}$.

Testing of hypothesis or confidence interval for $\hat{\beta}$ requires limiting distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{(i)}Y_{(i)}$.

Lemma

$$\Phi(\hat{\alpha}, \hat{\beta}) = S_{YY}(1 - \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\}).$$

$$\hat{\beta}^2 = \max\{r^2_{X\pi_1(Y)}, r^2_{X\pi_2(Y)}\} S_{YY}/S_{XX}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(i)}$ and $\frac{1}{n}\sum_{i=1}^{n} X_{(i)}Y_{(n-i+1)}$ are used in computing $\hat{\beta}$. Their limits lead to that of $\hat{\beta}$, and would aid in the estimation of the bias of $\hat{\beta}$.
Testing of hypothesis or confidence interval for $\hat{\beta}$ requires limiting distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{(i)}Y_{(i)}$.

General results for functions of marginal order statistics will be discussed that would aid in establishing

$$\hat{\beta} \xrightarrow{a.e} \beta_1 = sign(\beta)\sqrt{\beta^2 + \sigma_\epsilon^2 \sigma_X^{-1}}$$

Strong Law

$\{(X_i^{(1)}, \ldots, X_i^{(k)}), i = 1, 2, \ldots\}$ random vectors
j-th marginals $\{X_1^{(j)}, X_2^{(j)}, \ldots,\}$ are i.i.d. $\sim F_j$
$X_{n:i}^{(j)}$ is the i-th order statistic of $\{X_1^{(j)}, \ldots, X_n^{(j)}\}$.

ϕ measurable function on $\mathbb{R}^k$

$$\gamma(u) = \phi(F_1^{-1}(u), \ldots, F_k^{-1}(u)),\ 0 < u < 1.$$

Theorem

If F_j are continuous, and ϕ satisfies some smoothness conditions and well behaved near the tails of F_j, then

$$\frac{1}{n} \sum_{i=1}^{n} \phi(X_{n:i}^{(1)}, \ldots, X_{n:i}^{(k)}) \xrightarrow{a.e.} \int_0^1 \gamma(y) dy$$

Assumptions for Strong Law

1. The function $\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))$ is continuous at where $u_1 = u, \ldots, u_k = u, 0 < u < 1$.
   This implies γ is continuous.

Assumptions for Strong Law

1. The function $\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))$ is continuous at where $u_1 = u, \ldots, u_k = u, 0 < u < 1$.
   This implies γ is continuous.
2. For some A and $c_0 > 0$,

$$|\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))| \leq A \left(1 + \sum_{j=1}^{k} |\gamma(u_j)|\right),$$

whenever $(u_1, \ldots, u_k) \in (0, c_0)^k \cup (1 - c_0, 1)^k$

Assumptions for Strong Law

1. The function $\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))$ is continuous at where $u_1 = u, \ldots, u_k = u, 0 < u < 1$.
   This implies γ is continuous.
2. For some A and $c_0 > 0$,

$$|\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))| \leq A \left(1 + \sum_{j=1}^{k} |\gamma(u_j)| \right),$$

   whenever $(u_1, \ldots, u_k) \in (0, c_0)^k \cup (1 - c_0, 1)^k$
3. γ is integrable on $(0, 1)$.

Additional assumptions for CLT

4. $\lim_{u \to 0^+} \sqrt{u}(|\gamma(u)| + |\gamma(1-u)|) = 0$

Additional assumptions for CLT

4. $\lim_{u \to 0^+} \sqrt{u}(|\gamma(u)| + |\gamma(1-u)|) = 0$
5. $|\gamma|$ and ψ_j^2 are integrable, where for $0 < u < 1$

$$\psi_j(u) = \frac{\partial \phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))}{\partial u_j}\Big|_{(u,\ldots,u)}$$

Additional assumptions for CLT

4. $\lim_{u\to 0^+} \sqrt{u}(|\gamma(u)| + |\gamma(1-u)|) = 0$
5. $|\gamma|$ and ψ_j^2 are integrable, where for $0 < u < 1$

$$\psi_j(u) = \frac{\partial \phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))}{\partial u_j}\Big|_{(u,\ldots,u)}$$

6. ψ_j and $\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))$ are smooth functions.

Additional assumptions for CLT

4. $\lim_{u \to 0^+} \sqrt{u}(|\gamma(u)| + |\gamma(1-u)|) = 0$
5. $|\gamma|$ and ψ_j^2 are integrable, where for $0 < u < 1$

$$\psi_j(u) = \frac{\partial \phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))}{\partial u_j}\Big|_{(u,\ldots,u)}$$

6. ψ_j and $\phi(F_1^{-1}(u_1), \ldots, F_k^{-1}(u_k))$ are smooth functions.

Example:

$$\phi(F_1^{-1}(u), F_2^{-1}(v)) = \min(u,v)^{-\alpha}(1 - \max(u,v))^{-\alpha}$$

$$\gamma(u) = u^{-\alpha}(1-u)^{-\alpha}, \qquad 0 < \alpha < \frac{1}{2}$$

CLT

Theorem

Let $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(k)})$ *be i.i.d. Assume for any pair* $(1 \leq j \neq r \leq k)$, *the joint distribution* $F_{j,r}$ *of* $(X_i^{(j)}, X_i^{(r)})$ *is continuous. Under regularity assumptions,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\phi(X_{n:i}^{(1)}, \ldots, X_{n:i}^{(k)}) - \int_0^1 \gamma(y)dy) \xrightarrow{dist} N(o, \sigma^2)$$

CLT

Theorem

Let $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(k)})$ *be i.i.d. Assume for any pair* $(1 \le j \ne r \le k)$*, the joint distribution* $F_{j,r}$ *of* $(X_i^{(j)}, X_i^{(r)})$ *is continuous. Under regularity assumptions,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\phi(X_{n:i}^{(1)}, \ldots, X_{n:i}^{(k)}) - \int_0^1 \gamma(y)dy\right) \xrightarrow{dist} N(o, \sigma^2)$$

$$\begin{aligned}\sigma^2 =& 2 \sum_{1 \le j \ne r \le k} \int_0^1 \int_0^1 [F_{j,r}(F_j^{-1}(x), F_r^{-1}(y)) - xy] \psi_j(x)\psi_j(y)dxdy \\ &+ 2 \sum_{j=1}^{k} \int_0^1 \int_0^y x(1-y)\psi_j(x)\psi_j(y)dxdy\end{aligned}$$

Example 1

X_i, i.i.d. $\sim F$ $\quad Y_i$, i.i.d. $\sim G$.
F, G continuous

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_{n:i} - Y_{n:i})^2$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_i^2 + Y_i^2) - \lim_{n\to\infty} \frac{2}{n} \sum_{i=1}^{n} X_{n:i} Y_{n:i}$$

$$= E(X_1^2) + E(Y_1^2) - 2 \int_0^1 F^{-1}(u) G^{-1}(u) du$$

$$= E(X_1^2) + E(Y_1^2) - 2E(X_1 G^{-1}(F(X_1)))$$

Example 2

$(U_1, \ldots, U_k), (X_1^{(1)}, \ldots, X_1^{(k)}), (X_1^{(1)}, \ldots, X_1^{(k)}), \ldots$ i.i.d.
The marginals U_j are uniformly distributed
$F_{j,r}$ is the joint distribution of (U_j, U_r).
$\phi(u_1, \ldots, u_k) = u_1^{a_1} \cdots u_k^{a_k},\ a_j \geq 1$.

$$\frac{1}{n} \sum_{i=1}^{n} (X_{n:i}^{(1)})^{a_1} \cdots (X_{n:i}^{(k)})^{a_k} \xrightarrow{a.e.} \frac{1}{a_1 + \cdots + a_k + 1}$$

Example 2

$(U_1, \ldots, U_k), (X_1^{(1)}, \ldots, X_1^{(k)}), (X_1^{(1)}, \ldots, X_1^{(k)}), \ldots$ i.i.d.
The marginals U_j are uniformly distributed
$F_{j,r}$ is the joint distribution of (U_j, U_r).
$\phi(u_1, \ldots, u_k) = u_1^{a_1} \cdots u_k^{a_k},\ a_j \geq 1$.

$$\frac{1}{n} \sum_{i=1}^{n} (X_{n:i}^{(1)})^{a_1} \cdots (X_{n:i}^{(k)})^{a_k} \xrightarrow{a.e.} \frac{1}{a_1 + \cdots + a_k + 1}$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_{n:i}^{(1)})^{a_1} \cdots (X_{n:i}^{(k)})^{a_k} - \frac{1}{a_1 + \cdots + a_k + 1} \right] \xrightarrow{dist.} N(0, \sigma^2)$$

Example 2

$(U_1, \ldots, U_k), (X_1^{(1)}, \ldots, X_1^{(k)}), (X_1^{(1)}, \ldots, X_1^{(k)}), \ldots$ i.i.d.
The marginals U_j are uniformly distributed
$F_{j,r}$ is the joint distribution of (U_j, U_r).
$\phi(u_1, \ldots, u_k) = u_1^{a_1} \cdots u_k^{a_k}, \ a_j \geq 1$.

$$\frac{1}{n} \sum_{i=1}^{n} (X_{n:i}^{(1)})^{a_1} \cdots (X_{n:i}^{(k)})^{a_k} \xrightarrow{a.e.} \frac{1}{a_1 + \cdots + a_k + 1}$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(X_{n:i}^{(1)})^{a_1} \cdots (X_{n:i}^{(k)})^{a_k} - \frac{1}{a_1 + \cdots + a_k + 1} \right] \xrightarrow{dist.} N(0, \sigma^2)$$

$$\sigma^2 = 2 \sum_{1 \leq j < r \leq k} a_j a_r E(U_j U_r)^M + \frac{(2M-3)(M^2-2)}{2M+1} \sum_{j=1}^{k} a_j^2$$

$M = a_1 + \cdots + a_k$

Back to regression

$\{(X_i, Y_i), 1 \le i \le n\}$ i.i.d. bivariate normal with correlation ρ, means μ_1, μ_2, and standard deviations σ_1, σ_2.

$$X_1 \sim F, \quad Y_1 \sim G, \quad G^{-1}(F(x)) = \mu_2 + \frac{\sigma_2}{\sigma_1}(x - \mu)$$

Strong Law

$$\frac{1}{n}\sum_{i=1}^{n} X_{n:i} Y_{n:i} \xrightarrow{a.e.} E(X_1 G^{-1}(F(X_1))) = \mu_1\mu_2 + \sigma_1\sigma_2$$

Weak convergence

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{n:i}Y_{n:i} - \mu_1\mu_2 - \sigma_1\sigma_2) \xrightarrow{dist.} N(0, \sigma^2)$$

$$\sigma^2 = \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + (1+\rho^2)\sigma_1^2\sigma_2^2 + 2\rho\,\mu_1\mu_2\,\sigma_1\sigma_2$$

Back to regression

$\{(X_i, Y_i), 1 \le i \le n\}$ i.i.d. bivariate normal with correlation ρ, means μ_1, μ_2, and standard deviations σ_1, σ_2.

$$X_1 \sim F, \quad Y_1 \sim G, \quad G^{-1}(F(x)) = \mu_2 + \frac{\sigma_2}{\sigma_1}(x - \mu)$$

Strong Law

$$\frac{1}{n}\sum_{i=1}^{n} X_{n:i}Y_{n:i} \xrightarrow{a.e.} E(X_1 G^{-1}(F(X_1))) = \mu_1\mu_2 + \sigma_1\sigma_2$$

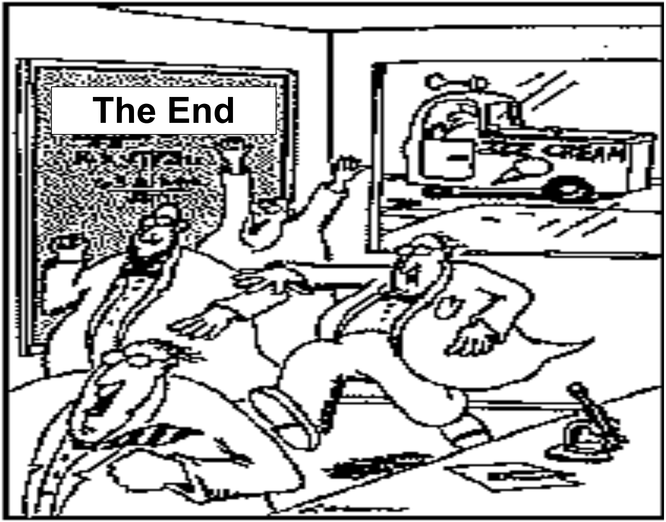
Weak convergence

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{n:i}Y_{n:i} - \mu_1\mu_2 - \sigma_1\sigma_2) \xrightarrow{dist.} N(0, \sigma^2)$$

$$\sigma^2 = \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + (1+\rho^2)\sigma_1^2\sigma_2^2 + 2\rho\,\mu_1\mu_2\,\sigma_1\sigma_2$$

– *Collaborators*

Zhidong Bai, Kwok-Pui Choi, and Vasudevan Mangalam


The End
ICE CREAM