Some largest eigenvalue problems in statistics.

Iain Johnstone, Stat + Biostat, UC Berkeley
(on leave, Stanford)

AGENDA

1. EXAMPLES

   - Principal Components Analysis

   - Canonical Correlations

2. GENERALITIES

   - One & Two Wishart problems

   - 'large' $n, p, (q)$ asymptotics

3. LARGEST EIGENVALUE(S) in null cases

   - for CCA = Jacobi ($\mathbb{R}$ & $\mathbb{C}$)

     [w. P. Forrester, in progress]

   - Liouville-Green

   - message for stat: 'large' = quite small

4. SOME OPEN QUESTIONS

REVIEW OF PCA : Goal: Dimensionality reduction

E.g: 1. Hotelling (1933)

$n = 140$ cases

$p = 4$ variables

reading < speed, power

arithmetic < speed, power

2. Craddock – Flood (1969)

$n = 1095$ Daily data, 500 mb heights

$p = 130$ Locations

→ Data Y_{ij} or $Y(t_i, x_j)$

i/t_i cases/times $i = 1, \ldots, n$

j/x_j variables/places $j = 1, \ldots, p$ $(\leq n)$

Mean centered $Y_{ij} \leftarrow Y_{ij} - \bar{Y}_{\cdot j}$ $\bar{Y}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} Y_{ij}$

'Covariance' matrix $S = (S_{jj'}) = Y'Y$

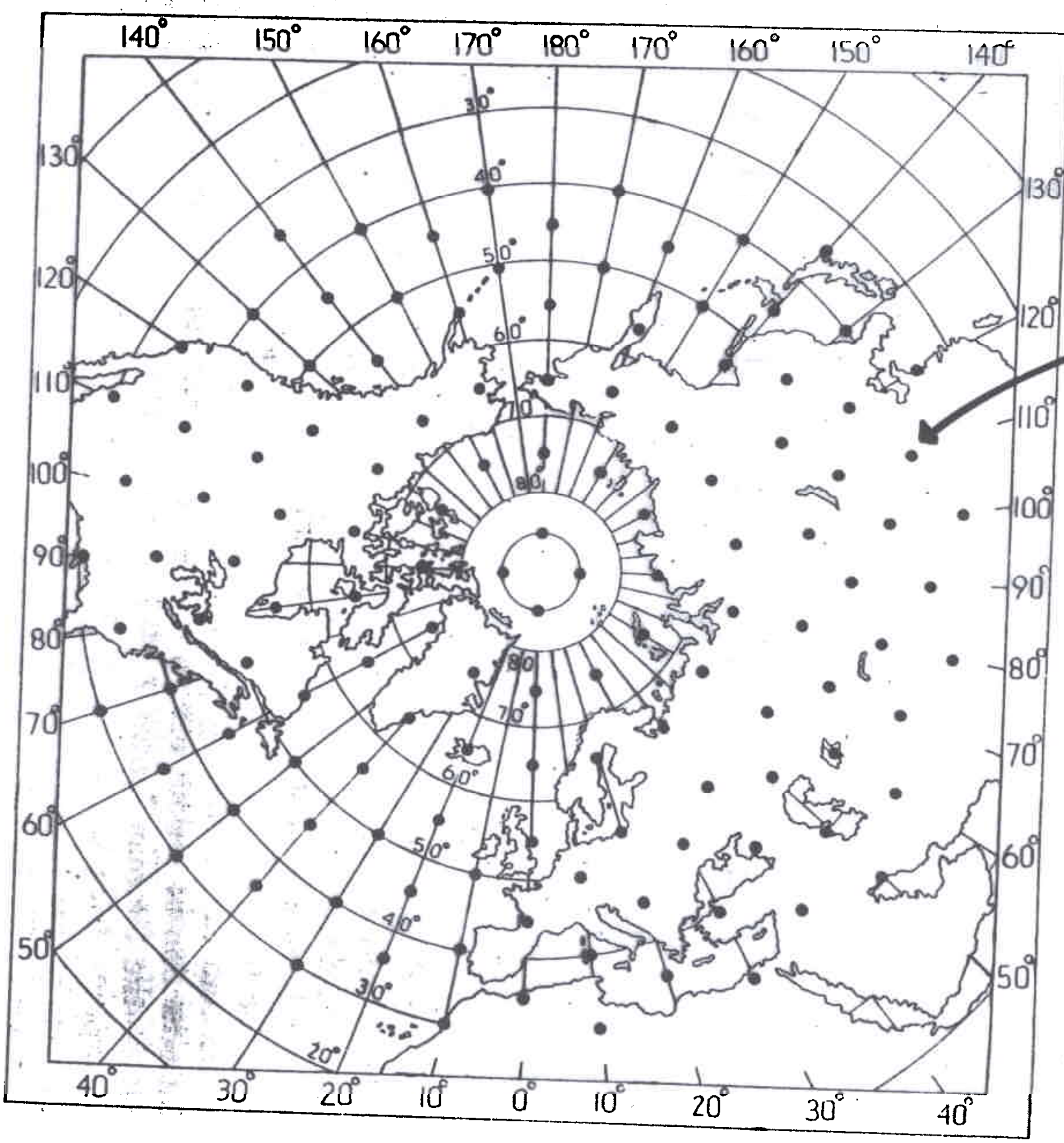
$$S_{jj'} = \sum_i Y_{ij} Y_{ij'}$$

Quart. J. R. Met. Soc. (1969), **95**, pp. 576-593

551.547.5 : 5

Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere

By J. M. CRADDOCK and C. R. FLOOD
Meteorological Office, Bracknell

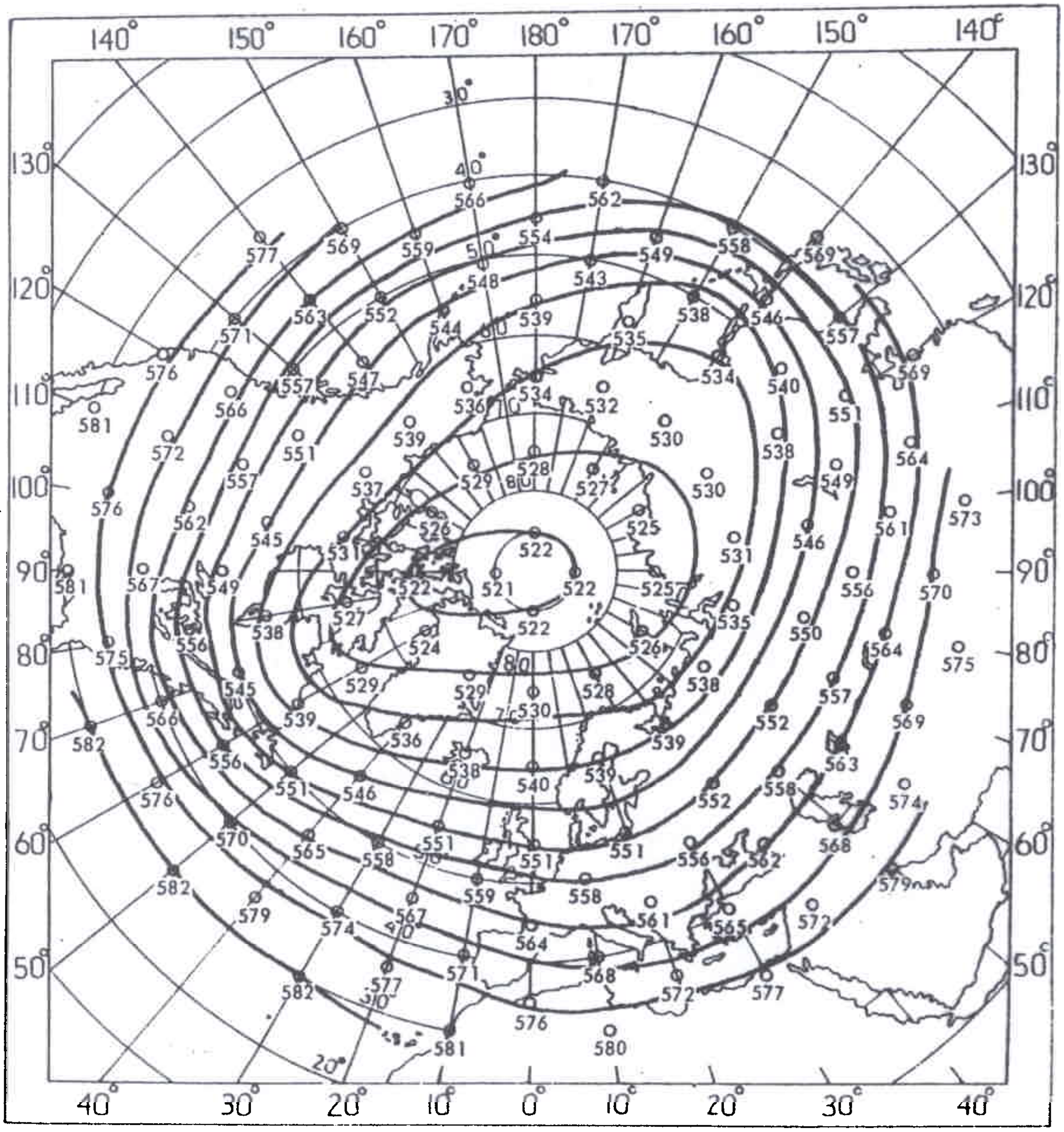


locations x_j

$p=$

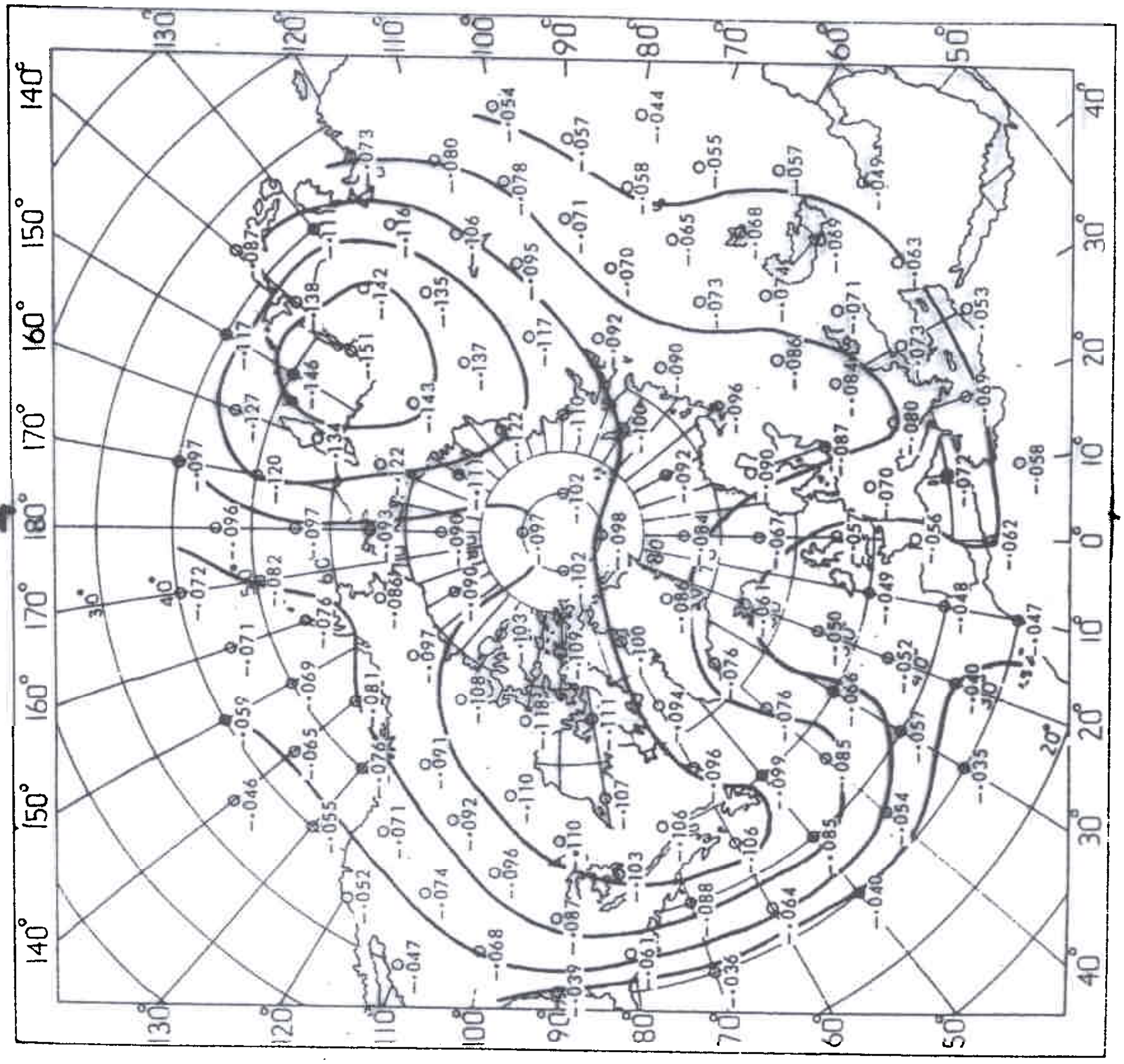
$j=1,\dots,130$

Mean 500mb field over $t_1 \ldots t_{1095}$

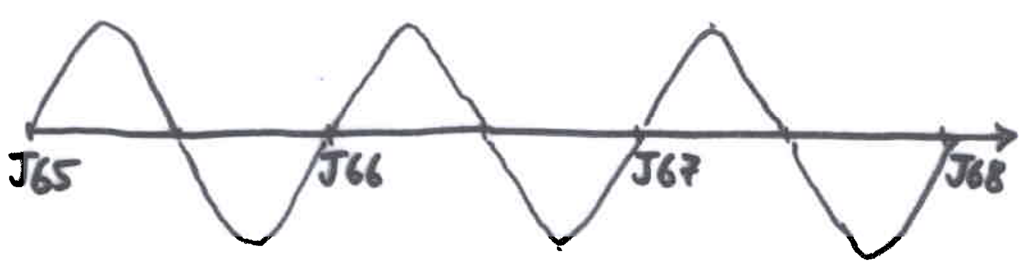


1st eigenvector $v_1(x_j)$ (e.o.f.)

$l_1 = 58.6\%$



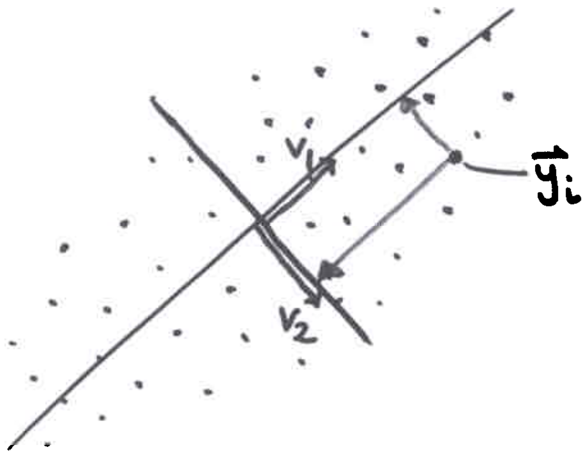
1st p.c. $z_1(t_i)$ $i = 1, \ldots, 1095$



MAX VARIANCE DIRECTION

n points in
p dimensions

$p = 2$



$$Y = \begin{bmatrix} \vec{y}_1 \\ \vdots \\ \vec{y}_n \end{bmatrix}$$

$$\mathrm{Var}\{ v'\vec{y}_i \} = v' S v$$

First p.c. variance + eigenvector

$$l_1 = \max \{ v' S v \; : \; v'v = 1 \}$$

v_1 = 1^{st} eigenvector / p.c. loadings

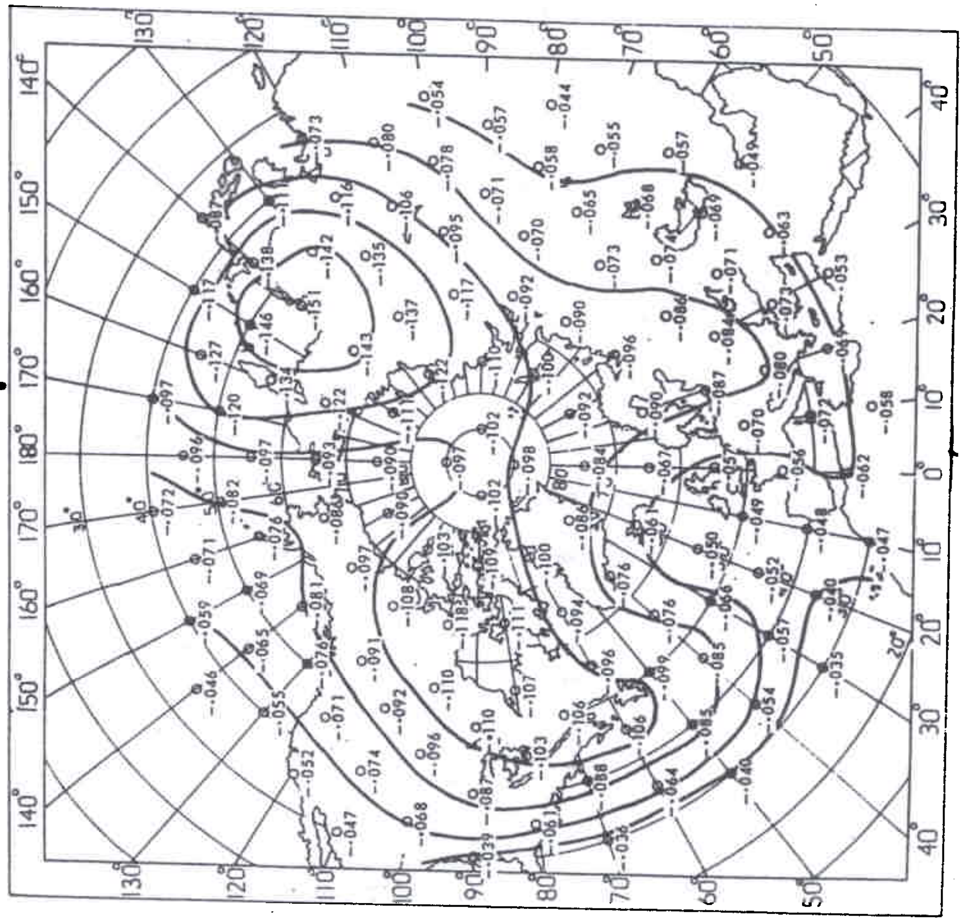
First p.c. :

$$z_1 = \sqrt{l_1}\, u_1 = Y v_1$$

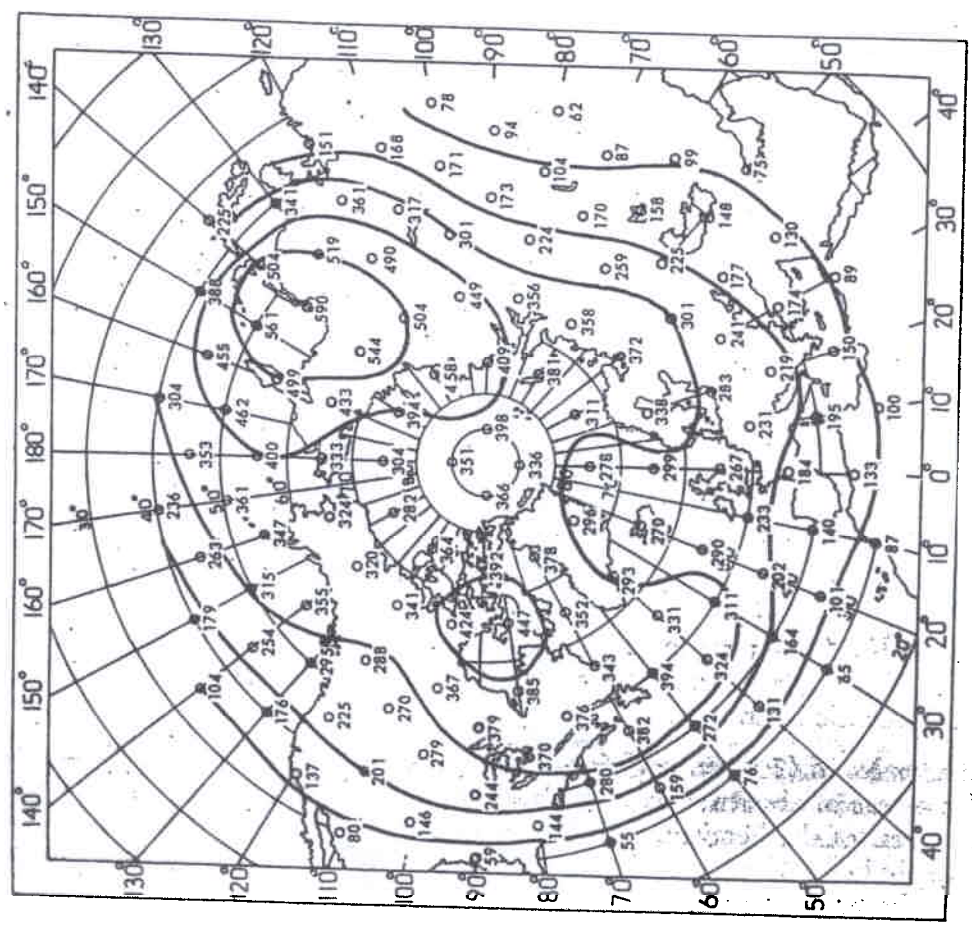
i.e. $\quad z_1(t_i) = \sum_j Y(t_i, x_j)\, v_1(x_j)$

Seasonal nature of 1st p.c.

1st eigenvector



Variances by location



ORTHOGONAL MAX VARIANCE DIRECTIONS

Iterate on p.c. variances

$$l_k = \max \{ v' S v : \quad v'v = 1, \quad v' v_m = 0 \quad m < k \}$$

→ $v_k = k^{th}$ eigenvector

$$z_k = \sqrt{l_k}\, u_k = Y v_k = k^{th} \text{ p.c.}$$

→ Singular Value Decomposition of data:

$$Y = U L V' = \sum_k \sqrt{l_k}\, u_k v_k' \qquad l_1 \geq l_2 \geq \dots$$

→ Best (linear!) r term approx^n to data Y:

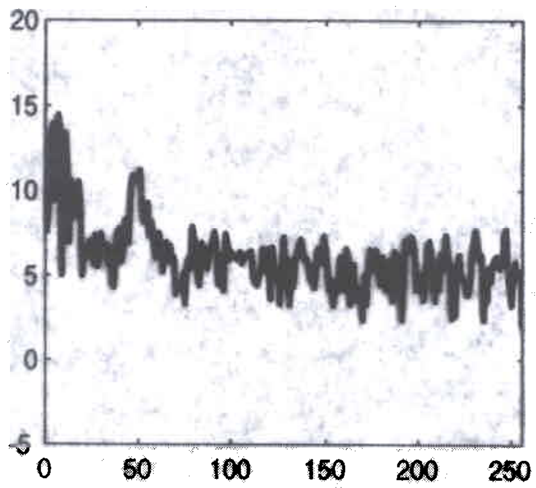
$$\hat{Y}_r(t_i, x_j) = \sum_{k=1}^{r} \sqrt{l_k}\, u_k(t_i)\, v_k(x_j)$$

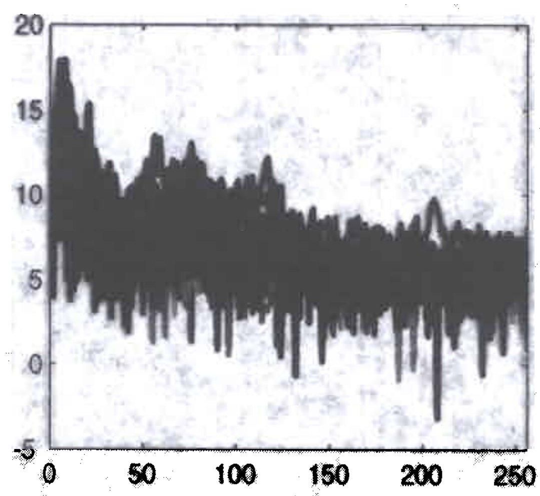
$$\| Y - \hat{Y}_r \|_F^2 = \sum_{r+1}^{p} l_k$$

→ wide use in data reduction/compression

- key role of eigenvalues $\{ l_k \}$

Example (via T. Hastie) Phoneme 'dcL'

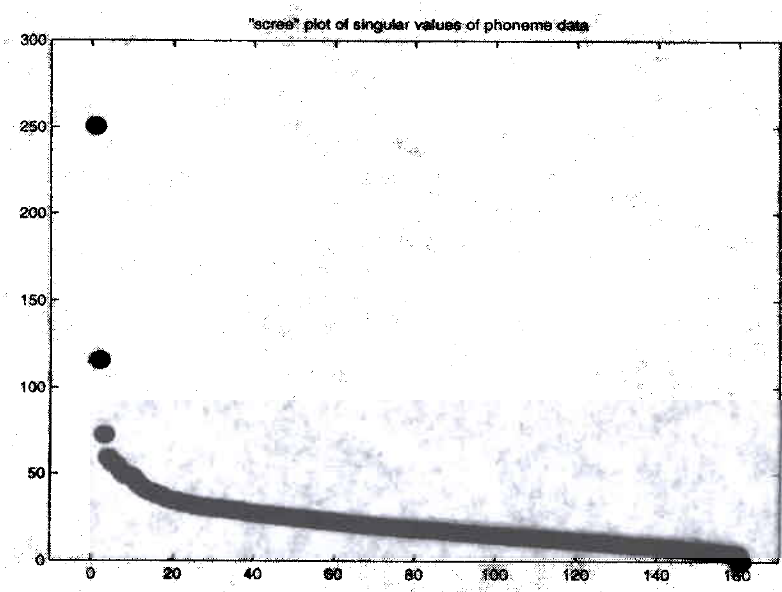




p = 256 log periodogram; n = 162 instances (50 spkrs)

GRAPHICAL METHODS : 'SCREE' PLOT

Cattell, 1966



CANONICAL CORRELATIONS ANALYSIS (CCA)

$$X = [\, x_1 \;\; x_2 \;\; \cdots \;\; x_p \,] \qquad Y = [\, y_1 \;\; y_2 \;\; \cdots \;\; y_q \,]$$

Find $a'x$ most correlated with $b'y$

$$\mathrm{Corr}(a'x,\, b'y) = \frac{a' S_{xy}\, b}{\sqrt{a' S_{xx}\, a}\,\sqrt{b' S_{yy}\, b}} \qquad \begin{array}{l} S_{xy} = X'Y \\ S_{xx} = X'X \\ \vdots \qquad \vdots \end{array}$$

$$r_k^2 = \max \left\{ a' S_{xy}\, b \;:\; \begin{array}{l} a' S_{xx}\, a = b' S_{yy}\, b = 1 \\ a' S_{xx}\, a_j = b' S_{yy}\, b_j = 0 \quad j < k \end{array} \right\}$$

→ determinantal equation

$$(*) \qquad \det\left(S_{xy} S_{yy}^{-1} S_{yx} - r^2 S_{xx} \right) = 0$$

→ $\quad r_1^2 \geq r_2^2 \geq \cdots \geq r_p^2 \qquad \left(\begin{array}{l} \text{and } a_1, \ldots, a_p \\ \qquad\; b_1, \ldots, b_p \end{array} \right)$

→ how many r_k^2 are "significant" ?

$$(*) \Leftrightarrow \qquad \det\left(A - r^2 (A+B) \right) = 0$$

$$A = S_{xy} S_{yy}^{-1} S_{yx} \qquad B = S_{xx} - A$$

1942, Econometrica

REGRESSIONS BETWEEN SETS OF VARIABLES

By Frederick V. Waugh, Agricultural Marketing Administration, Washington D.C.

Professor Hotelling's paper, "Relations between Two Sets of Variates,"[1] should be widely known and his method used by practical statisticians. Yet, few practical statisticians seem to know of the paper, and

X = "wheat characteristics" (raw)

x_1 = kernel texture

x_2 = test weight

x_3 = damaged kernels

x_4 = foreign material

x_5 = crude protein in wheat

Y = "flour characteristics" (finished)

y_1 = wheat per bbl. of flour

y_2 = ash in flour

y_3 = crude protein in flour

y_4 = gluten quality index

$a_1'x$ index of wheat quality

$b_1'y$ ... flour ...

GOAL: highly correlated grading of raw + finished products

$p = 5 \qquad q = 4 \qquad n = 138$

Monthly Weather Review

Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis

T. P. BARNETT

Climate Research Group, Scripps Institution of Oceanography, La Jolla, CA 92093

R. PREISENDORFER

Pacific Marine Environment Laboratory, National Oceanographic and Atmospheric Administration, Seattle, WA 98115

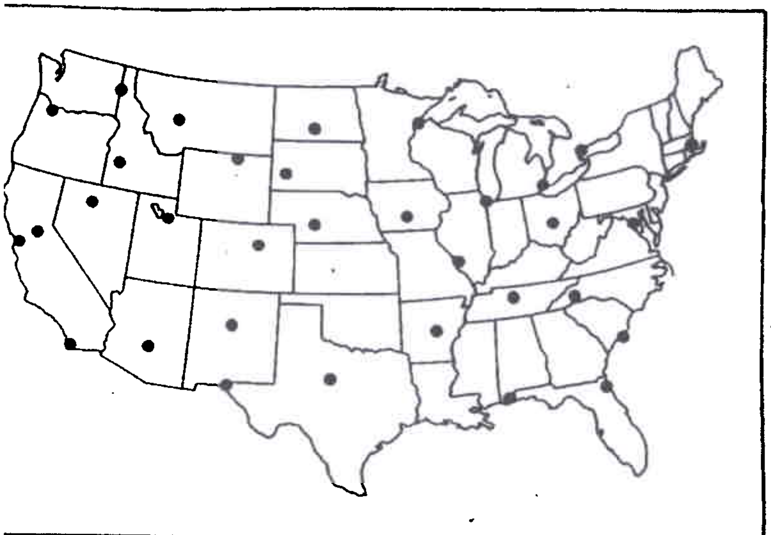


FIG. 1. Locations of stations/districts providing surface air temperature predictand data.

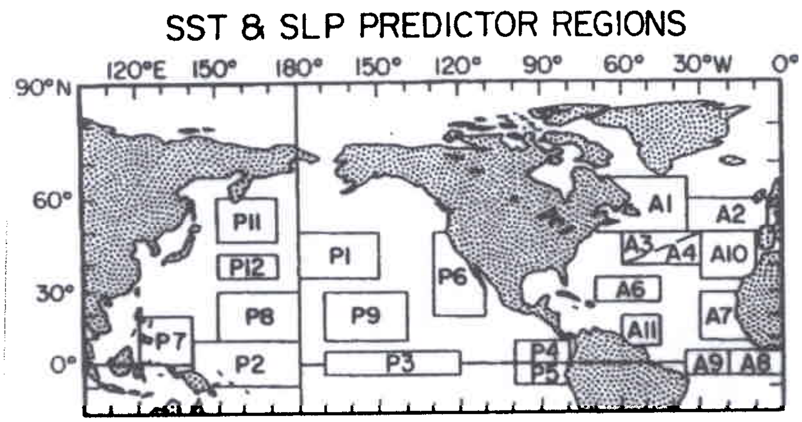


FIG. 2. SST from the large averaging areas shown above were used as predictor information. SLP predictor data came from the region 20°–70°N, 140°E to the Greenwich Meridian.

Y vars : surface air temps at 33 U.S. locations monthly data, 1931 – 1980

X vars : sea surface temp (SST) in 21 regions for 3 prior months in 2 seasons

$p = 126$ $q = 33$ $n = 600$

$(21 \times 3 \times 2)$ (50×12)

ROOTS OF A DETERMINANTAL EQUATION

$$|A - r^2(A+B)| = 0$$

where $\quad A = X'PX \qquad\qquad P = Y(Y'Y)^{-1}Y'$

$\qquad\quad\; B = X'P^{\perp}X \qquad\quad P^{\perp} = I - P$

Introduce probability: $\underset{n \times (p+q)}{[X : Y]} \sim N(0, I_n \otimes \Sigma) \qquad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$

<u>Null</u> distribution $\quad \Sigma_{xy} = 0 \quad \leftrightarrow \quad \rho_1 = \cdots = \rho_p = 0$

(and transform to $\Sigma_{xx} = I$)

BASIC SETTING : 2 INDEPENDENT WISHARTS

$$\left.\begin{aligned} A &\sim W_p(q, I) \\ B &\sim W_p(n-q, I) \end{aligned}\right\rangle \text{ indep., Wishart}$$

Multivariate Beta roots $:= (u_i)_{i=1}^p$

$$|u(A+B) - A| = 0$$

$\Leftrightarrow$ Multivariate F roots $:= (w_i)_1^p \qquad w = u/(1-u)$

$$|wB - A| = 0$$

MULTIVARIATE LINEAR MODEL

$$\underset{n\times p}{Y} = \underset{n\times q}{X}\ \underset{q\times p}{\beta} + \underset{n\times p}{E} \qquad E \sim N(0, I_n \otimes \Sigma)$$

q predictors $\qquad$ p responses

Let P = orthoprojection on column span of X

$$\begin{aligned} A &= Y'PY && \text{model SS} && \sim W_p(q, \Sigma, \Omega) \\ B &= Y'P^{\perp}Y && \text{error SS} && \sim W_p(n-q, \Sigma) \end{aligned} \Big\} \text{ inde}$$

Invariant tests of $H_0: \beta = 0$ $\qquad$ $(p \le q,\ p \le n-q$

use latent roots of AB^{-1}

E.g. Chemometrics (Skagerberg et.al. 92 in Breiman Friedman 9

$p = 6$ $\qquad$ output characteristics of polymers (correlated)

$q = 22$ $\qquad$ predictors: temps, feed rate

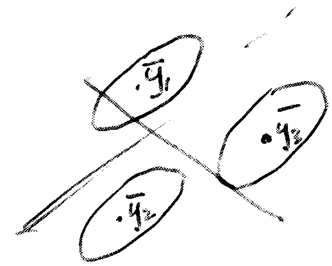
$n = 56$ $\qquad$ observations

MULTIPLE DISCRIMINANT ANALYSIS

$i = 1, \dots, q$ populations

n_i observations / pop'n on p variables

$$\vec{y}_{i1}, \dots, \vec{y}_{in_i} \sim N_p(\mu_i, \Sigma)$$



Between class:

$$A = \sum_i n_i(\bar{y}_i - \bar{y})(y_i - \bar{y})' \sim W_p(q-1, \Sigma, \Omega)$$

Within class:

$$B = \sum_{i,j} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)' \sim W_p(n-q, \Sigma)$$

Discriminant functions l_i solve

$$(A - w_i B) l_i = 0$$

- How many l_i are useful ($\leftrightarrow$ w_i signif)
- under H_0, roots of AB^{-1}

2. GENERALITIES: (non-central) WISHART

$$\underset{n\times p}{Y} = \begin{bmatrix} \vec{y}_1 + i\vec{y}_1' \\ \vdots \quad \vdots \\ \vec{y}_n + i\vec{y}_n' \end{bmatrix} \qquad \vec{y}_i + i\vec{y}_i' \overset{ind}{\sim} \mathbb{C}N_p(\vec{\mu}_i, \Sigma) \qquad M = \begin{bmatrix} \vec{\mu}_1 \\ \vdots \\ \vec{\mu}_n \end{bmatrix}$$

$$\underset{p\times p}{A} = Y^T Y \sim \mathbb{C}W_p(n, \Sigma; \Omega) \qquad \Omega = \Sigma^{-1} M^T M$$

- $M = 0 \rightarrow$ central Wishart
- $p = 1 \rightarrow \chi^2_n(\omega) \qquad \omega = M^T M / \sigma^2$

Roots of determinantal equations

"Wishart" $\det[A - l I] = 0$

— principal components, MDS

$B \sim W_p(n', \Sigma')$ independent of A:

"M. Beta" $\det[A - u(A+B)] = 0$

— canonical correlations

"M. F" $\det[A - wB] = 0$

— multivariate linear model

— mult discriminant; equality of covariances

Role of Eigenvalues in Some classical questions

Hypothesis testing: e.g. $H_0: \Sigma = I$ or $\Sigma_X = \Sigma_Y$

$H_0: M = 0$

$H_0: \mathrm{Cov}(X, Y) = 0$

$\rightarrow$ group invariant tests based on (l_i), (u_i), (w_i)

Estimation: e.g. of Σ; using

'mle': $\hat{\Sigma} = n^{-1} A = U L U^T$ $\quad U \in O(p)$, $L = \mathrm{diag}(l_i)$

Need to <u>shrink</u> (l_i):

e.g. $n = p = 10$ BUT $(l_i) = \underline{3.07}$ 1.40 1.12 0.78 0.51

$\Sigma = I$ (one sample!) 0.30 0.16 .095 .036 $\underline{.00}$

$\Rightarrow$ $\hat{\Sigma} = U \varphi(L) U^T$ Stein, Perlman-Lin, Haff, Berger, Loh, ...

$\rightarrow$ difficult to analyze for fixed n, p.

Classification, Regression, Clustering, ...

Non-null densities E.g. (l_i) from $A \sim W_p(n, \Sigma)$

$$f(l) = c\,|\Sigma|^{-\frac{n}{2}}\; {}_0\tilde{F}_0(-\tfrac{1}{2}\Sigma^{-1}, A) \prod_i l_i^{\frac{n-m-1}{2}} \prod_{i<j} (l_i - l_j)^{2-1}$$

$${}_0F_0(-\Sigma^{-1}, A) = \int_{\substack{U(p)\\ O(p)}} \operatorname{etr}\{-\Sigma^{-1} U A U^*\}\, dU$$

[$\Sigma = I \rightarrow$ Laguerre {U,O} E]

James, 1964 Hypergeom. functions of matrix argument

${}_0F_0$	exponential	χ^2, Wishart, $\lambda(A)$
${}_1F_0$	binomial	F, $\lambda(AB^{-1})$
${}_0F_1$	Bessel	non-central χ^2, Wishart, mea
${}_1F_1$	confluent hyper.	non-central F, $\lambda(AB^{-1})$
${}_2F_1$	Gaussian hyper.	multiple R^2, CCA

- influence on practice ?
- useful approximations via RMT ?

TRADITIONAL ASYMPTOTICS p fixed, $n \to \infty$

$$A = X^T X = U L U^T \sim W_p(n, \Sigma) \qquad \Sigma = \mathrm{diag}(\lambda_1 > \cdots > \lambda_r > 1 = \lambda_{r+1} = \cdots = \lambda_p)$$

$$\sqrt{\tfrac{n}{2}} \left(\tfrac{L}{n} - \lambda \right) \xrightarrow{\mathcal{D}} D \qquad \text{Girshick, 39} \to \text{Anderson, 63}$$

$$\begin{cases} D_1, \ldots, D_r \overset{ind}{\sim} N(0, \lambda_i^2) \\ D' = (D_{r+1}, \ldots, D_p) \sim GOE(p-r) \end{cases}$$

$$f_{D'}(d) = c\, e^{-\frac{1}{2}\sum_{r+1}^{p} d_j^2} \prod_{r<j<k\le p} |d_j - d_k|$$

Null case: $r = 0$; → not (traditionally) useful

Large n, p (+ q) asymptotics:

	Craddock Flood	Phoneme	Barnett-Preisendorfer	Waugh	Chemometrics
n	1095	161	600	137	56
p	130	256	126	4	6
q			33	5	22

First try: $p/n \to c_1 > 0$, $q/n \to c_2 > 0$

3. 'CASE STUDY' : **Largest eigenvalues in null cases:**

[PCA etc.] $\quad A = X^T X \sim \mathbb{C}W_p(n, I) \rightarrow l_1 \geq l_2 \geq \cdots \geq l_p$

$$f(l) = c \prod_i e^{-\frac{1}{2} l_i} \, l_i^{\frac{n-p-1}{2}} \prod_{i<j} (l_i - l_j)^{2-1}$$

Laguerre $\{U, O\}$ E

Empirical spectrum: Marčenko-Pastur + ...

Strong law for l_1 : Geman, + ... Bai (99)

Tracy-Widom limit for rescaled l_1 : Johansson 2000 $\mathbb{C}$

IMJ 2001 $\mathbb{R}$.

[CCA etc.] $\quad A \sim W_p(q, I)$, $B \sim W_p(n-q, I)$ } indep, $\quad p \leq q, n-q$

$u_1 \geq u_2 \geq \cdots \geq u_p$ solve $\det[A - u(A+B)] = 0$

$$f(u) = c \prod_i u_i^{\frac{1}{2}(q-p-1)} (1-u_i)^{\frac{1}{2}(n-q-p-1)} \prod_{i<j} (u_i - u_j)^{2-1}$$

Jacobi $\{U, O\}$ E

Fisher, Girshick, Hsu, Mood, Roy 1939

LIMITING EMPIRICAL SPECTRUM

Wachter 1980

$$A \sim W_p(q, I) \perp\!\!\!\perp B \sim W_p(n-q, I) \qquad p \le q, n-q$$

$$|A - u_i(A+B)| = 0 \qquad u_1 \ge u_2 \ge \cdots \ge u_p \ge 0$$

NOTE: $u_i = r_i^2$ squared correlation scale

ASSUME: $0 < \sin^2 \frac{\gamma_0}{2} \leftarrow \frac{p - \frac{1}{2}}{n-1} \le \frac{q - \frac{1}{2}}{n-1} \rightarrow \sin^2 \frac{\phi_0}{2}$

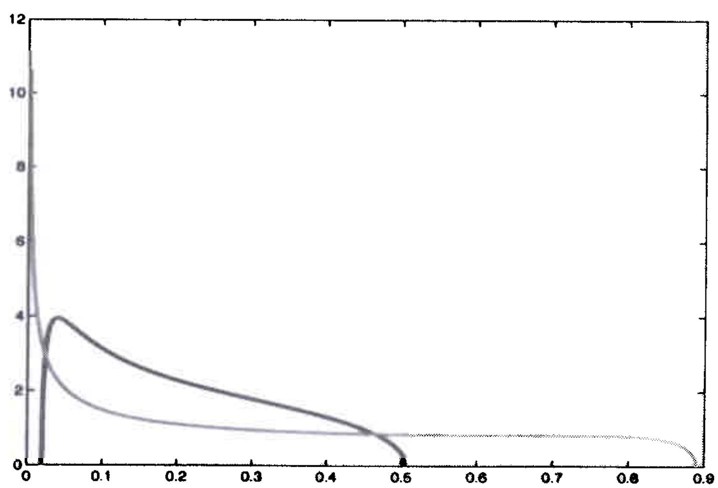
THEN: $F_p(u) = p^{-1} \# \{ i : u_i \le u \} \longrightarrow \int_0^u f(u')\, du'$

$$f(u) = \frac{c_u \sqrt{(\mu_+ - u)(u - \mu_-)}}{u(1-u)} \qquad c_u = 2\pi \sin^2 \frac{\gamma_0}{2}$$

where

$$\mu_\pm = \cos^2\left(\frac{\pi}{2} - \frac{\phi_0 \pm \gamma_0}{2}\right)$$

$f(u)$



u

$\uparrow$ μ_-

$\uparrow$ $\mu_+ \doteq (.7)^2$

$\uparrow$

$n = 200, \quad p = 20$
$q = 40$

"evil RA"

c.f. also

Bai - Yin - Krishnaiah
Silverstein ('85)
Bai (99)

LIMITING DISTRIBUTION FOR u_1

$$A \sim W_p(q, I) \perp\!\!\!\perp B \sim W_p(n-q, I)$$ (resp complex) Wishart

$$|A - u_i(A+B)| = 0 \qquad u_1 \geq u_2 \geq \cdots \geq u_p \qquad p \leq q, n-q$$

Let $p, q(p), n(p) \to \infty$

Set $$\sin^2 \frac{\gamma_p}{2} = \frac{p - 1/2}{n-1} \qquad \sin^2 \frac{\varphi_p}{2} = \frac{q - 1/2}{n-1}$$

$$\mu_{p+} = \cos^2\left(\frac{\pi}{2} - \frac{\varphi_p + \gamma_p}{2}\right)$$

$$\sigma_{p+}^3 = \frac{1}{(2n-2)^2} \frac{\sin^4(\varphi_p + \gamma_p)}{\sin \varphi_p \cdot \sin \gamma_p}$$

Assume $(\gamma_p, \varphi_p) \to (\gamma_0, \varphi_0)$ with $0 < \gamma_0 \leq \varphi_0$

Theorem $$\frac{u_1 - \mu_{p+}}{\sigma_{p+}} \xrightarrow{\mathcal{D}} W_1 \quad (\text{resp } W_2)$$

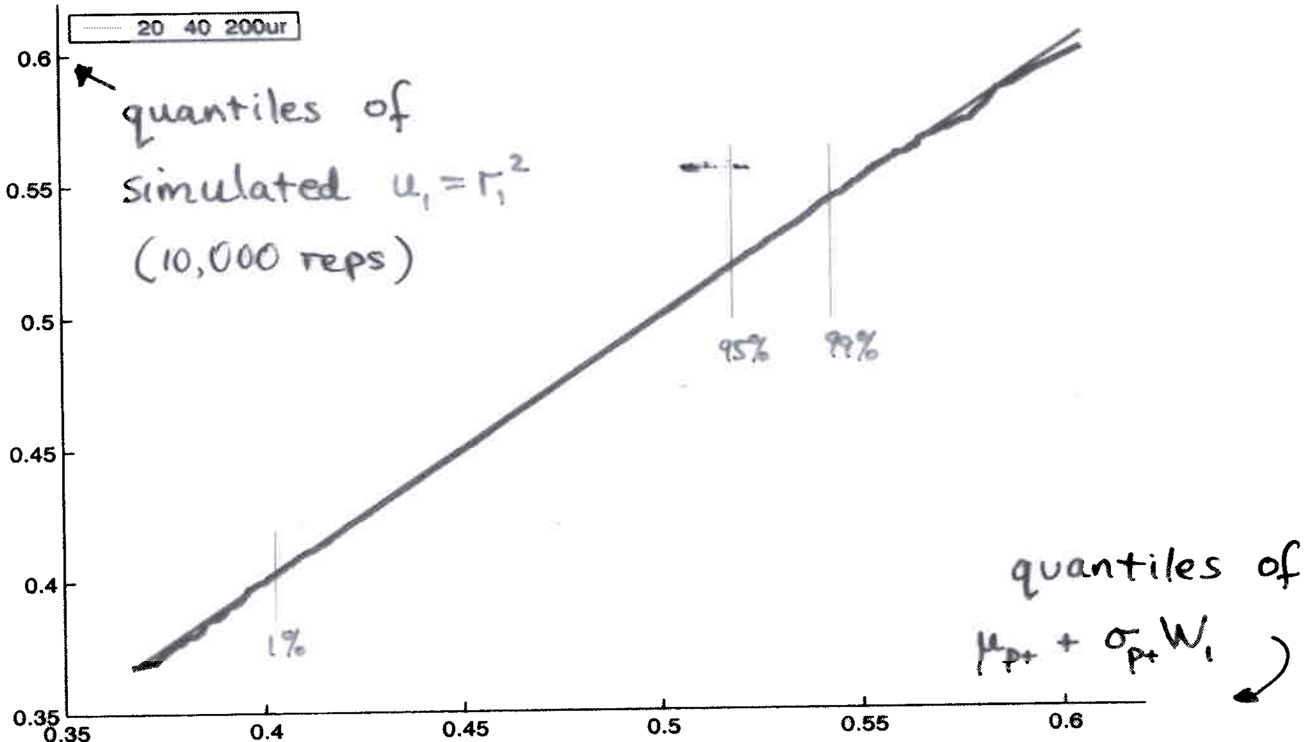
(Tracy–Widom laws)

|| Real case: $p - \frac{1}{2}$, $q - \frac{1}{2}$, $n-1$ v. important for numerics

– reduces error from $O(p^{-1/3})$ to $O(p^{-2/3})$

Complex case: tweak $\mu_{p+}^c, \sigma_{p+}^c$ to get $O(p^{-2/3})$

Finite (p,q,n) simulations of $u_1 = r_1^2$



Percentile for W_1	TW (μ, σ)	$\boxed{20,40,200}$ (.494, .024)	5,10,50 (.478, .061)	p, q, n 2,4,20 (.442, .117)	2 * SE
-3.90	.01	.010	.008	.000	(.002)
-3.18	.05	.052	.049	.009	(.004)
-2.78	.10	.104	.099	.046	(.006)
-1.91	.30	.311	.304	.267	(.009)
-1.27	.50	.507	.506	.498	(.010)
-0.59	.70	.706	.705	.711	(.009)
0.45	.90	**.904**	**.910**	**.911**	(.006)
0.98	.95	**.950**	**.955**	**.958**	(.004)
2.02	.99	**.990**	**.992**	**.995**	(.002)

$\hat{P}\left[\frac{u_1 - \mu_p}{\sigma_p} \le 0.45 \;\middle|\; (n,p,q) = (20,40,200)\right]$

REMARKS

- $p^{-2/3}$ scale of variability for u_1

- 95 % tile $\doteq$ $\mu_{p+} + \sigma_{p+}$

  99 % tile $\doteq$ $\mu_{p+} + 2\sigma_{p+}$

- if $\mu_{p+} > .8$ better approx on logit scale:

$$v_i = \log u_i/(1-u_i)$$

$$\mu_{v+} = \log \frac{\mu_{p+}}{1-\mu_{p+}} \qquad \sigma_{v+} = v'(\mu_{p+})\,\sigma_{p+} = \frac{\sigma_{p+}}{\mu_{p+}(1-\mu_{p+})}$$

- <u>Smallest</u> eigenvalue : previous assumptions <u>and</u> $\gamma_0 < \varphi_0$

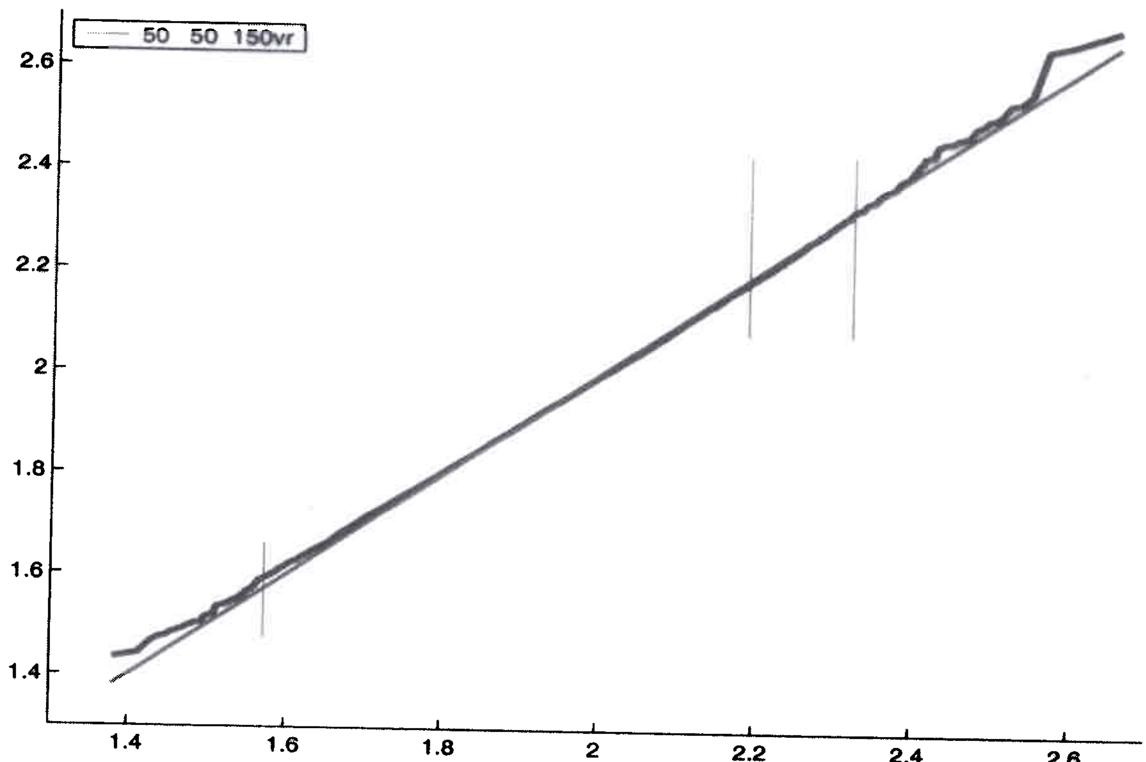
$$\sigma_{p-}^3 = \frac{1}{(2n-2)^2} \frac{\sin^4(\varphi_p - \gamma_p)}{\sin\varphi_p \cdot \sin\gamma_p}$$

then $\dfrac{\mu_{p-} - u_p}{\sigma_{p-}} \xrightarrow{\mathcal{D}} W_1 \quad (W_2)$

- Corresponding limit dist^{ns} for $u_2 \geq \cdots \geq u_k$

  $u_{p-k} \geq \cdots \geq u_{p-1}$

  k <u>fixed</u>.



Percentile	TW (μ, σ)	50,50,150 (2.06, .127)	5,5,15 (1.93, .594)	p, q, n, 2,2,6 (1.69, 1.11)	2 * SE
-3.90	.01	.007	.002	.010	(.002)
-3.18	.05	.042	.023	.037	(.004)
-2.78	.10	.084	.062	.074	(.006)
-1.91	.30	.289	.262	.264	(.009)
-1.27	.50	.499	.495	.500	(.010)
-0.59	.70	.708	.725	.730	(.009)
0.45	.90	.905	**.919**	**.931**	(.006)
0.98	.95	.953	**.959**	**.966**	(.004)
2.02	.99	.990	**.991**	**.993**	(.002)

QUALITY OF APPROXIMATION: AN ANALOGY

	t-statistic $\sqrt{n}\bar{x}/s$	largest root l_1 of S
Model	$x_i \overset{ind}{\sim} N(\mu, \sigma^2)$ $H_0: \mu = 0$	$\vec{X}_i \sim N(0, \Sigma)$ $H_0: \Sigma = I$
Exact Law	$t \sim t_{n-1}$	$l_1 \sim LOE(p, n-p)$
Approx Law	$\Phi(x) = \int_{-\infty}^{x} \phi(s)\,ds$	$F_1(x) = \exp\{-\frac{1}{2}\int_x^\infty q(s) + (x-s)^2 q(s)\, d$

Second-Order Accuracy

$$P\{t_{(n-1)} \le x\} = \Phi(x) + O\left(\left(\tfrac{1}{\sqrt{n}}\right)^2\right)$$

'Correlation' functions

$$\sigma_{np} S_p(\mu_{np} + \sigma_{np} x, \mu_{np} + \sigma_{np} y)$$

$$\to S(x, y) + O(p^{-2/3})$$

<u>Conj:</u> $P\{l_1 \le \mu_{np} + \sigma_{np} x\}$

$$= F_1(x) + O(p^{-2/3})$$

Robustness to Model

- analogous issues
- Soshnikov (2001)

Convergence Strategy for Extreme Eigenvalues

$$f(x_1, \dots, x_N) = \frac{1}{N!} (q) \det \left[K_{N(1)} (x_i, x_j) \right]$$

Correlation functions:

$$\varrho_{N,k} (x_1, \dots, x_k) = (q) \det \left[K_{N(1)} (x_i, x_j) \right]$$

[Soshnikov] Convergence of moments $\int_{I_k} \varrho_{N,k} \to \int_{I_k} \varrho_k \quad \forall I_k$

(+ ...) implies, for determinantal point fields F_N, F

$$\mathcal{L}\{x_{(1)}, \dots, x_{(k)} \mid F_N\} \xrightarrow[N \to \infty]{} \mathcal{L}\{x_{(1)}, \dots, x_{(k)} \mid F\} \qquad \underline{k \text{ fixed}}$$

so, need (uniform) convergence of rescaled kernel

$$\sigma_N K_N(\mu_N + \sigma_N s, \mu_N + \sigma_N t) \to K_A(s,t)$$

AND to reduce error from $O(N^{-1/3})$ to $O(N^{-2/3})$

need good approx^{ns} to center μ_N

scale σ_N

JACOBI $\to$ AIRY To show

$$\sigma_N K_N(\mu_N+\sigma_N s, \mu_N+\sigma_N t) \to K_A(s,t) = \frac{1}{s-t}\begin{vmatrix} Ai(s) & Ai(t) \\ Ai'(s) & Ai'(t) \end{vmatrix}$$

use Christoffel - Darboux property

$$K_N(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\,\phi_k(y) = \frac{a_N}{x-y}\begin{vmatrix} \varphi_N(x) & \varphi_N(y) \\ \varphi_{N-1}(x) & \varphi_{N-1}(y) \end{vmatrix}$$

So, need to show $\boxed{\zeta N^{-1/6}\varphi_N(\mu_N+\sigma_N s) \simeq Ai(s)} + O(N^{-2/3})$

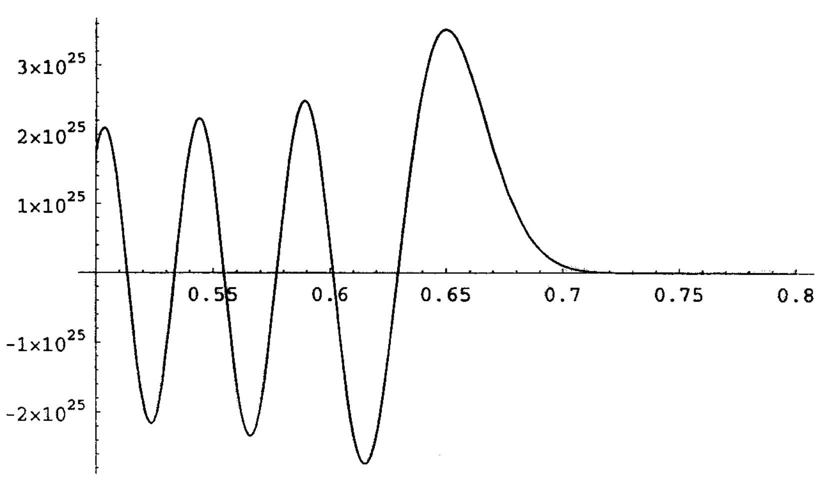
where $\varphi_N(x) = h_N^{-1/2}(1-x)^{\alpha/2}(1+x)^{\beta/2}P_N^{\alpha,\beta}(x)$,

and, treble asymptotics: $\begin{pmatrix} \alpha \\ \beta \\ N \end{pmatrix} = \begin{pmatrix} n-p-q \\ q-p \\ p \end{pmatrix} \to \infty$

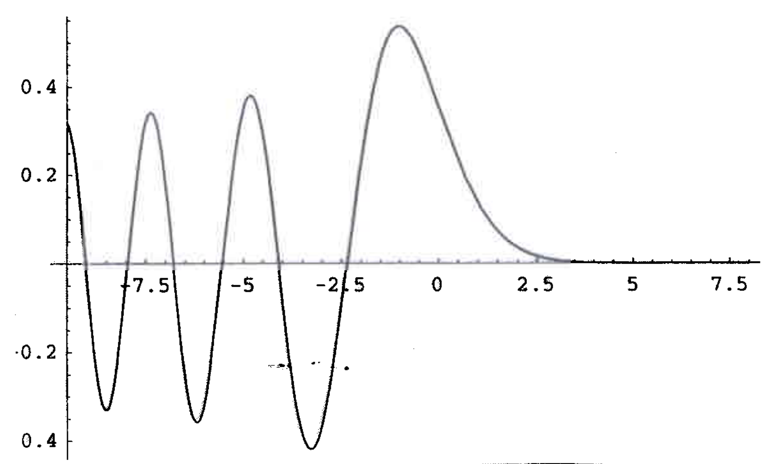
<u>Example</u>:

$(\alpha, \beta, N) = (200, 100, 100) \iff (p, q, n) = (100, 200, 500)$

$(1-x)^{\alpha/2}(1+x)^{\beta/2}P_N^{\alpha,\beta}(x)$



$Ai(s)$



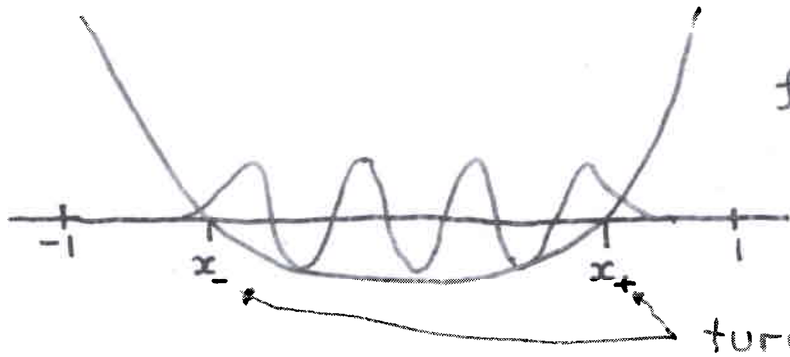
$\boxed{Ai''(s) = s\,Ai(s)}$

LIOUVILLE - GREEN asymptotics for ortho polynomials

e.g. Laguerre $L_N^{\alpha}(x)$, Jacobi $P_N^{\alpha,\beta}(x)$ α,β,N large

Satisfy 2nd order o.d.e.

$$(*) \qquad \frac{d^2 w}{dx^2} = \{\kappa^2 f(x) + g(x)\} w(x) \qquad \kappa = \kappa(N, \alpha, \beta) \text{ large}$$



$$f(x) = \frac{(x - x_-)(x - x_+)}{4(1-x^2)^2}$$

Transform $\quad \dfrac{d^2 W}{d^2 \zeta} = [\kappa^2 \zeta + \psi(\zeta)]\, W(\zeta)$

$$\zeta^{3/2}(x) = \int_{x_+}^{x} \sqrt{f}\,(y)\, dy$$

ERROR BOUND (Olver, 74) $\exists$ solution of $(*)$ with

$$w(x,\kappa) = \hat{f}^{-1/4}(x) \left\{ Ai(\kappa^{2/3}\zeta) + \varepsilon(x,\kappa) \right\} \qquad \hat{f}(x) = \frac{f(x)}{\zeta}$$

$$|\varepsilon(x,\kappa)| \leq |Ai|(\kappa^{2/3}\zeta) \left\{ \exp\left[\frac{c_0}{\kappa} V(\zeta)\right] - 1 \right\}$$

so long as $\quad V(\zeta) = \displaystyle\int_{\zeta}^{\infty} |\psi(v)|\, v^{-1/2}\, dv < \infty$

Virtues of Error Bound

- constrains approx^n : $V(\zeta)$ convergent

$$\Rightarrow \quad g(x) \sim \frac{-1/4}{(1-x)^2} \quad \Rightarrow \quad \kappa_N = 2N + \alpha + \beta + 1$$

$\Rightarrow$ uniform error bounds in x : $|\varepsilon(x,\kappa)| = O(1/N)$

Centering + Scaling constants: $x_N = x_+$

σ_N by matching: $\kappa^{2/3}\,\zeta(x_N + \sigma_N s) \simeq s$ $\quad s$ small

$$\Rightarrow \quad \sigma_N = \frac{1}{\kappa^{2/3}\,\dot{\zeta}(x_N)} \quad \Rightarrow \quad \sigma_N^3 = \frac{2\sin^4(\varphi+\gamma)}{\kappa^2 \sin\varphi \,\sin\gamma}$$

$\Rightarrow$ Jacobi $\rightarrow$ Airy approx^n: with $x = x_N + \sigma_N s$

$$\phi_N(x) = \left(\frac{\sigma_N \kappa_N}{1-x^2}\right)^{1/2} \left[Ai(s) + O(N^{-2/3}) \right]$$

Similarly for ϕ_{N-1} :

$$x_{N-1} = x(N-1, \alpha, \beta)$$
$$\sigma_{N-1} = \sigma(N-1, \alpha, \beta)$$

Lazy Higher Order Approximation Seek <u>good</u> μ, σ with

$$\sigma K_N(\mu + s\sigma, \mu + t\sigma) = K_A(s,t) + R_N \qquad (\dagger)$$

From Christoffel-Darboux,

$$K_N(x,y) = a_N \frac{\phi_N(x)\phi_{N-1}(y) - \phi_N(y)\phi_{N-1}(x)}{x-y}, \quad \text{with}$$

$$\phi_N(x) = \left(\frac{\sigma_N K_N}{1-x^2}\right)^{1/2}\left[Ai(s_N) + O(N^{-2/3})\right], \qquad s_N = \frac{x - x_N}{\sigma_N}$$

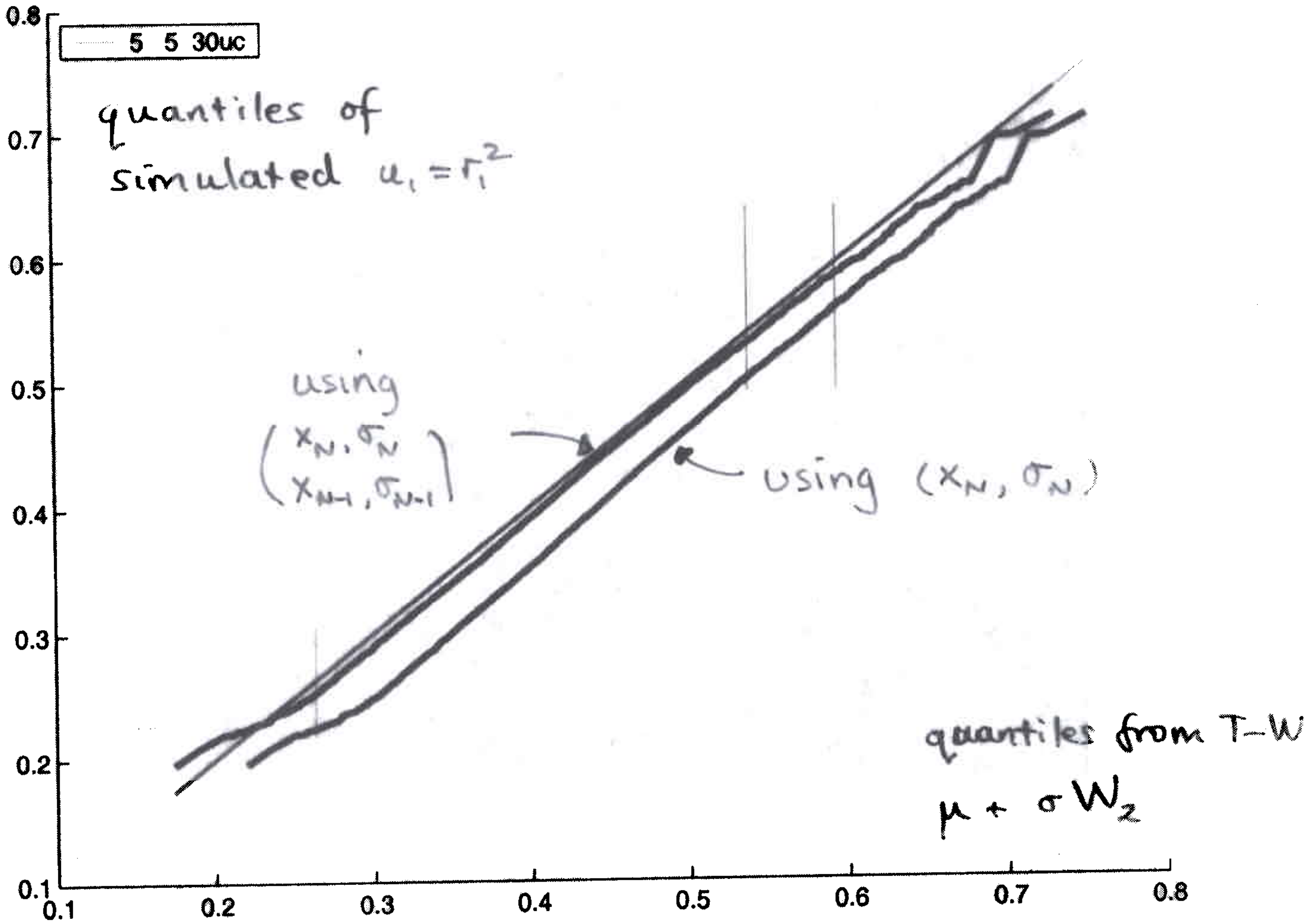
For ϕ_N use center x_N, scale σ_N,

but using (x_N, σ_N) in $(\dagger)$ $\Rightarrow R_N = O(N^{-1/3})$

For ϕ_{N-1} use center x_{N-1}, scale σ_{N-1}, $s_{N-1} = \dfrac{x - x_{N-1}}{\sigma_{N-1}}$

and with $s = \dfrac{s_N + s_{N-1}}{2}, \ldots$ $\Rightarrow \boxed{R_N = O(N^{-2/3})}$

$$\Rightarrow \quad \mu = \frac{x_N/\sigma_N + x_{N-1}/\sigma_{N-1}}{1/\sigma_N + 1/\sigma_{N-1}}, \qquad \frac{1}{\sigma} = \frac{1}{2}\left(\frac{1}{\sigma_N} + \frac{1}{\sigma_{N-1}}\right)$$



5 5 30uc
quantiles of simulated $u_1 = r_1^2$
using $\begin{pmatrix} x_N, \sigma_N \\ x_{N-1}, \sigma_{N-1} \end{pmatrix}$
using (x_N, σ_N)
quantiles from T-W
$\mu + \sigma W_2$
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

THE REAL CASE + higher order approximation

$$\rho_{N,k}(x_1, \dots, x_k) = \underset{1 \le i,j \le k}{\mathrm{q\,det}} \left[f_{N,1}(x_i, x_j) \right]$$

$$f_{N,1}(x,y) = \begin{bmatrix} K_{N,1}(x,y) & I_{N,1}(x,y) \\ D_{N,1}(x,y) & S_{N,1}(x,y) \end{bmatrix}$$

Laguerre OE : closed form representation in Widom, 99

Jacobi OE : use Adler - Forrester - Nagao - van Moerbeke, OC

$$K^{\alpha-1,\beta-1}_{N,1}(x,y) = \underbrace{\sqrt{\frac{1-x^2}{1-y^2}}\, K^{\alpha,\beta}_{N-1,2}(x,y)} + \underbrace{a_{N-1} \frac{k_{N-1}}{2} \psi_{N-1}(y)(\epsilon * \psi_{N-2})(x)}$$

$\psi_N(x) = \phi_N(x) / \sqrt{1-x^2}$

$\epsilon(x) = \frac{1}{2} \operatorname{sgn} x$

$x = x_{N-1} + s\,\sigma_{N-1}$

$y = \dots \; t \dots$

$$K_A(s,t) - \frac{\Delta_N}{2} Ai(s) Ai(t) + O(N^{-2/3})$$

$$\frac{1}{2} Ai(t) \int_{-\infty}^{s} Ai(u)\, du + \frac{\Delta_N}{2} Ai(s) Ai(t) + O(N^{-2/3})$$

- 'miraculous' cancellation of $O(N^{-1/3})$ terms

$\Rightarrow$ much improved small-(n,p,q) approx^{n}.

SOME OPEN QUESTIONS

E.g. 'Non-white' Wishart $\qquad A \sim \mathbb{C}W_p(n, \Sigma)$

$$f(l;\Sigma) = c|\Sigma|^{-n} \prod_i l_i^{n-p} \int_{U(p)} \mathrm{etr}(-\Sigma^{-1} U A U^*)\, dU\, \Delta^2(l)$$

(James)

$$= c \prod_i \lambda_i \prod_i l_i^{n-p} \frac{\det[e^{-l_k/\lambda_j}]\, \Delta(l)}{\Delta(\lambda_j^{-1})}$$

(HCIZ)

$\rightarrow$ determinantal correlation functions

Distribution of l_1 (or $l_2, \dots$) as $n/p \to \gamma$. ?

E.g. a) $\Sigma = (\lambda_1, \dots, \lambda_\eta > 1 = \lambda_{\eta+1} = \dots = \lambda_p)$ $\qquad \eta$ FIXED

$$\lambda_\eta \simeq n^{-1}(\sqrt{n} + \sqrt{p})^2$$

b) power law $\quad \lambda_k = c k^{-\beta}$

Progress :
- Jinho Baik on (a)
- S. Peché + G. Ben Arous

A Vaguer List:

- Extreme eigenvalues:
  - 'Bootstrap' C.I's + validity as $n/p \to \gamma$?
  - strong law behavior for general Σ

- Eigenvectors:
  - Consistency, asymptotic dist^n as $n/p \to \gamma$
  - Effect of regularization (FDA)
  - 'Sparse' versions of PCA (Lu)

- Empirical dist^n of eigenvalues
  - (further) statistical uses of Marčenko-Pastur
  - use of CLT for $\Sigma\, h(l_i)$

- Estimation of Large covariance matrices
  - sparsity for non-unit variances
  - Bayes estimates (Guionnet)

- Classification / Clustering
  - models for large covariance structures (Bickel-Levina)

- Issues from application domains
  - climate studies, face recognition, document retrieval ..