Random Matrices and Multivariate Statistical Analysis

Iain Johnstone, Statistics, Stanford

imj@stanford.edu

Agenda

- Classical multivariate techniques
  - Principal Component Analysis
  - Canonical Correlations
  - Multivariate Regression
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
  - Single Wishart
  - Double Wishart
- Largest Eigenvalue
  - Single Wishart
  - Double Wishart
- Concluding Remarks

Classical Multivariate Statistics

Canonical methods are based on spectral decompositions:

One matrix (Wishart)

- Principal Component analysis
- Factor analysis
- Multidimensional scaling

Two matrices(independent Wisharts)

- Multivariate Analysis of Variance (MANOVA)
- Multivarate regression analysis
- Discriminant analysis
- Canonical correlation analysis
- Tests of equality of covariance matrices

Gaussian data matrices

p variables

$$X = (\ x_{ik}) = \begin{bmatrix} \cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \end{bmatrix} \begin{matrix} n \\ \text{cases} \end{matrix} = \begin{bmatrix} \vec{x}_1^{\,T} \\ \vdots \\ \vec{x}_n^{\,T} \end{bmatrix} = \begin{bmatrix} \Big| & \Big| & \cdots & \Big| \end{bmatrix}$$

Independent rows: $\vec{x}_i \sim N_p(0, \Sigma), \qquad i = 1, \ldots n$

or: $X \sim N(0, I_n \otimes \Sigma_p)$

Zero mean $\Rightarrow$ no centering in **sample covariance matrix:**

$$S = (S_{kk'}), \qquad S = \frac{1}{n} X^T X, \qquad S_{kk'} = \frac{1}{n} \sum_{i=1}^{n} x_{ik} x_{ik'}$$

$$nS \sim W_p(n, \Sigma)$$

Principal Components Analysis

Hotelling, 1933 $\quad X_1, \ldots, X_n \sim N_p(\mu, \Sigma)$,

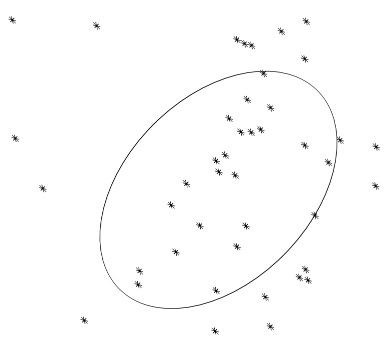
Low dim. subspace “explaining most variance”:

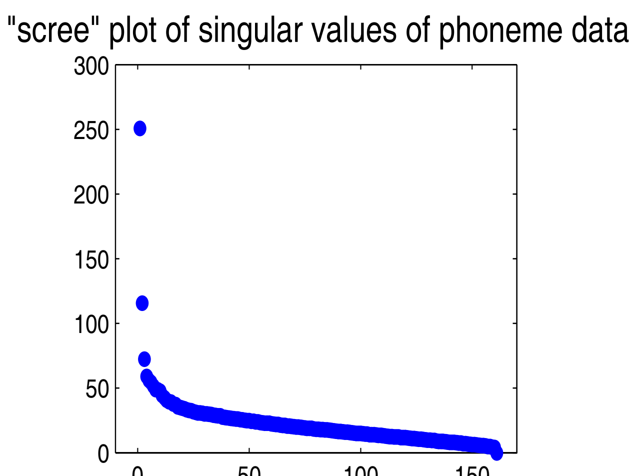
$$l_i = \max\{u'Su \;:\; u'u = 1, u'u_j = 0, j < i\}$$

Eigenvalues of Wishart: $\quad A = nS \sim W_p(n, \Sigma)$:

$$Au_i = l_i u_i \qquad l_1 \geq \ldots \geq l_p \geq 0.$$

Key question: How many l_i are “significant”?





Canonical Correlations

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \cdots \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$ jointly $p + q$ – variate normal.

"Most predictable criterion": (Hotelling, 1935, 1936).

$$\max_{u_i, v_i} \text{Corr}\,(u_i' X, v_i' Y)$$

$$\Rightarrow \quad A v_i = r_i^2 (A + B) v_i, \qquad r_1^2 \geq \ldots \geq r_p^2.$$

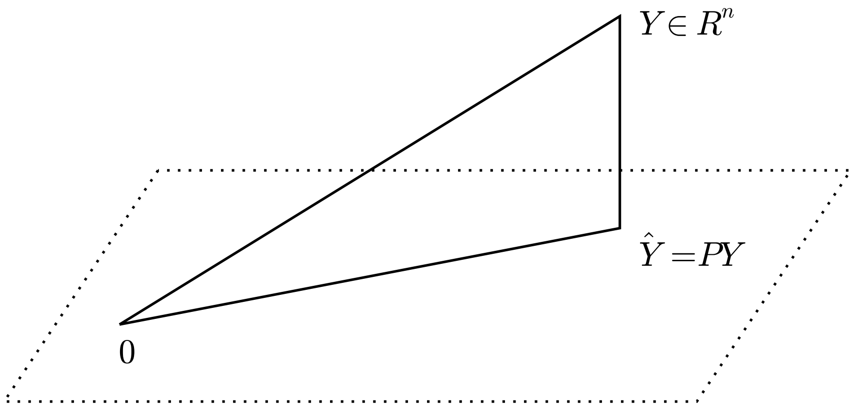
Two *independent* Wishart distributions:

$$A \sim W_p(q, \Sigma), \qquad B \sim W_p(n - q, \Sigma).$$

Multivariate Multiple Regression

$$\underset{n\times p}{Y} = \underset{n\times q}{X}\;\underset{q\times p}{B} + \underset{n\times p}{U} \qquad U \sim N_p(0, I\otimes\Sigma)$$

$n = \#$ observations; $p = \#$ response variables; $q = \#$ predictor variables



$P = X(X^TX)^{-1}X^T$ projection on span{cols(X)}

$$Y^TY = Y^TPY + Y^T(I-P)Y$$

H : hypothesis SSP $+$ E : error SSP

$H \sim W_p(q, \Sigma)$ indep of $E \sim W_p(n-q, \Sigma)$

Agenda

- Classical multivariate techniques
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
- Largest Eigenvalue
- Concluding Remarks

Hypothesis Testing

Null hypothesis H_0, nested within *Alternative* hypothesis H_A

Test Statistics: functions of *eigenvalues*: $T = T(l_1, \ldots, l_p)$.

Null hypothesis distribution: $P(T > t | H_0$ true).

RMT offers tools for

evaluation, and approximation based on $p \to \infty$

Single Wishart $A \sim W_p(n, I)$ e-vals $\det(A - l_i I) = 0$.
Test $H_0 : \Sigma = I$ (or λI) versus $H_A : \Sigma$ unrestricted.

Double Wishart $H \sim W_p(q, \Sigma), E \sim W_p(n - q, \Sigma)$
independently. Eigenvalues $\det(H - l_i(E + H)) = 0$.

Typical hypothesis test (e.g. from $Y = XB + U$):

$H_0 : B = 0$ versus $H_A : B$ unrestricted

Likelihood Ratio Test

If $X \sim N_p(0, I_p \otimes \Sigma)$, the density

$$f_\Sigma(X) = \det(2\pi\Sigma)^{-n/2} \exp\{-(n/2)\mathrm{tr}\Sigma^{-1}S\}$$

Log likelihood $\Sigma \to \ell(\Sigma|X) =$

$$\log f_\Sigma(X) = c_{np} - \tfrac{n}{2} \log\det\Sigma - \tfrac{n}{2}\mathrm{tr}\Sigma^{-1}S$$

Maximum likelihood occurs at $\hat{\Sigma} = S$:

$$\max_\Sigma \ell(\Sigma|X) = c_{np} - \tfrac{n}{2} \log\det S$$

Likelihood ratio test of $H_0 : \Sigma = I$ vs. $H_A : \Sigma$ unrestricted:

$$\log LR = \max_{\Sigma \in H_0} \ell(\Sigma|X) - \max_{\Sigma \in H_A} \ell(\Sigma|X)$$

$$= c_{np} + \tfrac{n}{2}\Big(\sum_i \log l_i - \sum_i l_i\Big)$$

Linear statistics in eigenvalues of S: $\sum_i \log l_i, \sum_i l_i$.

(Union-) Intersection Principle

Combine univariate test statistics:

$$H_0 : \Sigma = I \qquad \Leftrightarrow \qquad \bigcap_{|a|=1} H_{0a} : a^T \Sigma a = 1.$$

$\mathsf{Var}(a^T X) = a^T \Sigma a$, so reject H_{0a} if $\widehat{\mathsf{Var}}(a^T X) = a^T S a > c_a$

$$\begin{aligned} \text{Reject } H_0 \quad &\Leftrightarrow \text{reject } \textit{some } H_{0a} \\ &\Leftrightarrow \max_a a^T S a > c_{max} \\ &\Leftrightarrow l_{max}(S) > c_{max} \end{aligned}$$

Summary:
Likelihood ratio principle $\rightarrow$ linear statistics in eigenvalues
Intersection principle $\rightarrow$ extreme eigenvalues

Agenda

- Classical multivariate techniques
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
- Largest Eigenvalue
- Concluding Remarks

Eigenvalue densities - single Wishart

Statistics (n,p): $c\prod_{i=1}^{N} l_j^{\frac{n-p-1}{2}} e^{-l_i/2} \prod_{j<k} |l_j - l_k|$

Laguerre OE (N,α): $c\prod_{i=1}^{p} x_j^{\frac{\alpha}{2}} e^{-x_i/2} \prod_{j<k} |x_j - x_k|$

$$\begin{pmatrix} N \\ \alpha \end{pmatrix} \quad \leftrightarrow \quad \begin{pmatrix} p \\ n-p-1 \end{pmatrix}$$

$p =$ #variables

$n =$ sample size

Notation change has significance!

Statistics: no necessary relation between p and n;
traditional approximation uses p fixed, $n \to \infty$.

RMT: $N \to \infty$ with α fixed is most natural.
(in Stat, fixing $n-p$ would be less natural).

Eigenvalue densities - double Wishart

Statistics: If $H \sim W_p(q, I)$ and $E \sim W_p(n-q, I)$ are indep, then joint density of eigenvalues $\{u_i\}$ of $H(H+E)^{-1}$ is

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

With $$\begin{pmatrix} p \\ n-q-p \\ q-p \end{pmatrix} \leftrightarrow \begin{pmatrix} N+1 \\ \alpha \\ \beta \end{pmatrix}, \qquad u = (1+x)/2,$$

recover the Jacobi orthogonal ensemble

$$f(x) = c \prod_{i=1}^{N+1} (1-x_i)^{(\alpha-1)/2} (1+x_i)^{(\beta-1)/2} \prod_{i<j}^{N+1} |x_i - x_j|.$$

Convergence of Empirical Spectra

For e-values $\{l_i\}_{i=1}^{p}$ $\quad G_p(t) = p^{-1}\#\{l_i \le t\} \to G(t) = g(t)dt.$

Single Wishart (Marčenko-Pastur, 67) $A \sim W_p(n, I)$
If $p/n \to c > 0$,

$$g^{MP}(t) = \frac{\sqrt{(b_+ - t)(t - b_-)}}{2\pi ct}, \qquad b_\pm = (1 \pm \sqrt{c})^2.$$

Double Wishart (Wachter, 80) $\quad \det(H - l_i(H + E)) = 0.$
If $p \le q, \quad p/n \to c = \sin^2(\gamma/2) > 0, \quad q/n \to \sin^2(\phi/2),$

$$g^W(t) = \frac{\sqrt{(b_+ - t)(t - b_-)}}{2\pi ct(1-t)}, \qquad b_\pm = \sin^2(\frac{\phi \pm \gamma}{2}).$$

Agenda

- Classical multivariate techniques
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
  - Single Wishart
  - Double Wishart
- Largest Eigenvalue
- Concluding Remarks

Linear Statistics: Single Wishart

Approximate distributions:

Statistics: • Typically p *fixed*; standard χ^2 approximation,
• improvements by 'Bartlett correction'

RMT: • Central Limit Theorems (p large) for linear statistics of eigenvalues. Large literature

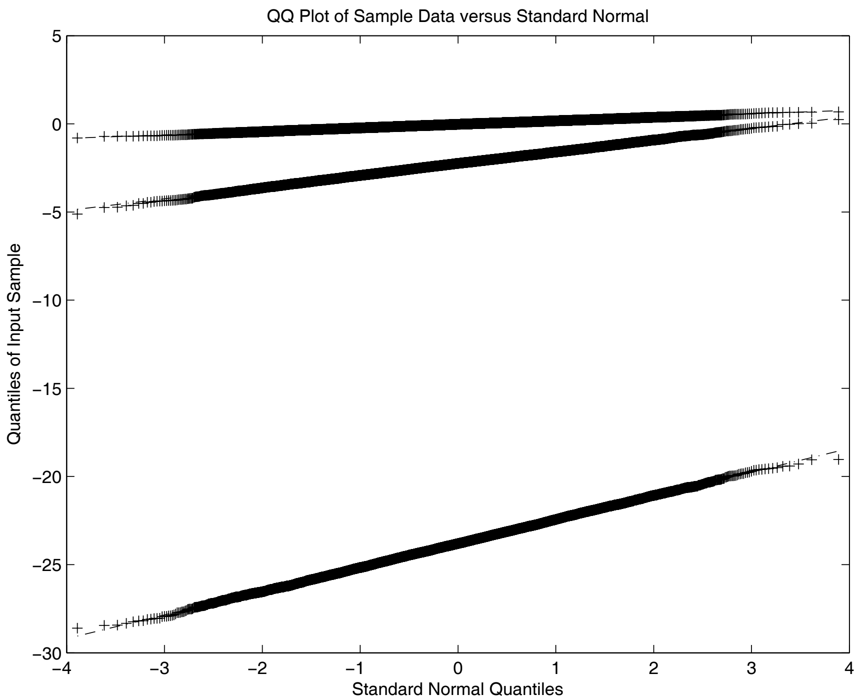
Jonsson (1982): $S \sim W_p(n, I)$, $p/n \to c > 0$ With
$d(c) = (1 - c^{-1}) \log(1 - c) - 1$,

$$\log\det S - pd(c) \xrightarrow{\mathcal{D}} N(\tfrac{1}{2}\log(1-c), -2\log(1-c)) \qquad (1)$$

$$\text{tr}S - p \xrightarrow{\mathcal{D}} N(0, 2c)$$

Surprise: quality of approximation in (1) for p small (e.g. 2!)

Small p asymptotics



n	p	qtile	pFix	pBaiS
100	2	0.90	0.923	0.899
100	20	0.90	1.000	0.900
100	60	0.90	1.000	0.902
1000	20	0.90	0.990	0.900
100	2	0.95	0.965	0.951
100	20	0.95	1.000	0.951
100	60	0.95	1.000	0.949
1000	20	0.95	0.997	0.950
100	2	0.99	0.995	0.992
100	20	0.99	1.000	0.990
100	60	0.99	1.000	0.990
1000	20	0.99	1.000	0.990

CLT for Likelihood Ratio distribution

Bai-Silverstein(2004)

$$\sum_{1}^{p} \mathbf{f}(l_i) - p \int \mathbf{f}(x) g^{MP}(x) dx \xrightarrow{\mathcal{D}} X_{\mathbf{f}} \sim N(EX_{\mathbf{f}}, \mathsf{Cov}(X_{\mathbf{f}})),$$

$$\mathsf{Cov}(X_f, X_g) = \frac{-1}{2\pi^2} \int_{\Gamma_1} \int_{\Gamma_2} \frac{f(z(m_1)) g(z(m_2))}{(m_1 - m_2)^2} dm_1 dm_2$$

$\Rightarrow$ CLT for null distribution of the LR test of $H_0 : \Sigma = I$,

$$\sum_{1}^{p} (\log l_i - l_i + 1) \xrightarrow{\mathcal{D}} N(pd(c) + \tfrac{1}{2} \log(1-c), 2[\log(1-c)^{-1} - c]).$$

Linear Statistics: Double Wishart

Hypothesis tests based on e-vals u_i of $H(H+E)^{-1}$, i.e.
e-vals $w_i = u_i/(1-u_i)$ of HE^{-1}.

Many standard tests are *linear statistics* $S_N(g) = \sum_1^p g(u_i)$:

- Wilks Λ: $\log \Lambda = \sum_1^p \log(1-u_i)$ [Likelihood ratio test]
- Pillai's trace $= \sum_1^p u_i$
- Hotelling-Lawley trace $= \sum u_i/(1-u_i) = \sum_1^p w_i$
- *Roy's largest root* $= u_{(1)}$.

Basor-Chen (05) *Unitary* case, formal; $N \to \infty$ α, β fixed.

$$S_N(g) - (2N + \alpha + \beta)a_g \xrightarrow{\mathcal{D}} N(0, b_{g \cdot g}),$$

$$[a_g = \frac{1}{2\pi}\int_{-1}^{1} \frac{g(x)}{\sqrt{1-x^2}}dx, \quad b_{g \cdot g} = \frac{1}{2\pi^2}\int_{-1}^{1} \frac{g(x)}{\sqrt{1-x^2}} P\int_{-1}^{1} \frac{\sqrt{1-y^2}}{y-x} g(y)dydx]$$

Agenda

- Classical multivariate techniques
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
- Largest Eigenvalue
  - Single Wishart
  - Double Wishart
- Concluding Remarks

Largest Eigenvalue - Single Wishart

'Usual' approach to maxima is (classically) infeasible:

$$\{l_{(1)} \leq x\} = \prod_{i=1}^{p} I\{l_i \leq x\}$$

Key role: *determinants*, not independence:

$$\prod_{i<j}(l_i - l_j) = \det[l_i^{k-1}]_{1\leq i,k\leq p}$$

$$\prod_{i=1}^{p} I\{l_i \leq x\} = \sum_{k=0}^{p}(-1)^k \binom{p}{k} \prod_{i=1}^{k} I\{l_i > x\}.$$

$$\cdots \Rightarrow P\{\max_{1\leq i\leq p} l_i \leq t\} = \sqrt{\det(I - K_p\chi_{[t,\infty)})}$$

$K_p(x, y)$ is (2×2 matrix) kernel uses {Laguerre, Jacobi} orthogonal polynomials via Christoffel-Darboux summation.

Tracy-Widom Limit

For real ($\beta = 1$, IMJ) or complex ($\beta = 2$, Johansson) data, if $n/p \to c \in (0, \infty)$:

$$F_p(s) = P\{l_1 \le \mu_{np} + \sigma_{np} s\} \to F_\beta(s),$$

with

$$\mu_{np} = (\sqrt{n} + \sqrt{p})^2, \qquad \sigma_{np} = (\sqrt{n} + \sqrt{p})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}\right)^{1/3}$$

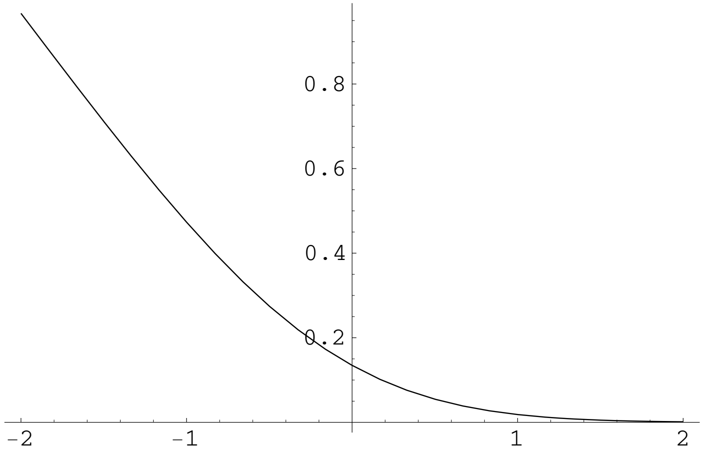
El Karoui (2004) In complex case, for refined μ'_{np}, σ'_{np},

$$|F_p(s) - F_2(s)| \le Ce^{-s} p^{-2/3}.$$

Also, results for
- $N \to \infty, p \to \infty$ separately, and
- under alternative hypotheses.

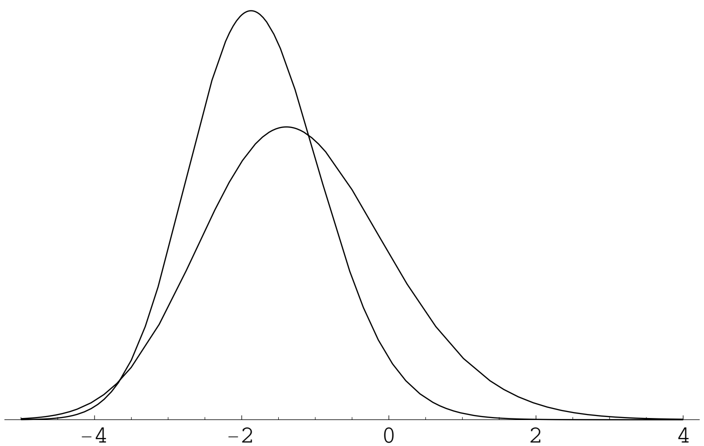
Painlevé II and Tracy-Widom

Painlevé II:



$$q'' = xq + 2q^3$$
$$q(x) \sim \mathrm{Ai}(x) \qquad \text{as} \quad x \to \infty$$

Tracy-Widom distributions:



$$F_2(s) =$$
$$\exp\{-\int_s^\infty (x-s)q^2(x)dx\}$$
$$F_1(s) =$$
$$(F_2(s))^{1/2}\exp\{-\tfrac{1}{2}\int_s^\infty q(x)dx\}.$$

Largest Root - Double Wishart

Assume $p, q(p), n(p) \to \infty$.

$$\frac{\gamma_p}{2} = \sin^{-1}\sqrt{\frac{p-\mathbf{.5}}{n-\mathbf{1}}}, \qquad \frac{\phi_p}{2} = \sin^{-1}\sqrt{\frac{q-\mathbf{.5}}{n-\mathbf{1}}}.$$

$$\mu_\pm = \cos^2\Big(\frac{\pi}{2} - \frac{\phi_p \pm \gamma_p}{2}\Big), \quad \sigma^3_{p+} = \frac{1}{(2n-\mathbf{2})^2}\frac{\sin^4(\phi_p+\gamma_p)}{\sin\phi_p \sin\gamma_p}.$$

Simply,

$$\frac{u_1 - \mu_+}{\sigma_+} \xrightarrow{\mathcal{D}} W_1 \sim F_1$$

More precisely, *logit transform* $\ell(u) = \log(u/(1-u))$: (IMJ,PJF)

$$|P\{\ell(u_1) \le \ell(\mu_+) + s\sigma_+\ell'(\mu_+)\} - F_1(s)| \le Ce^{-s/4}p^{-2/3}$$

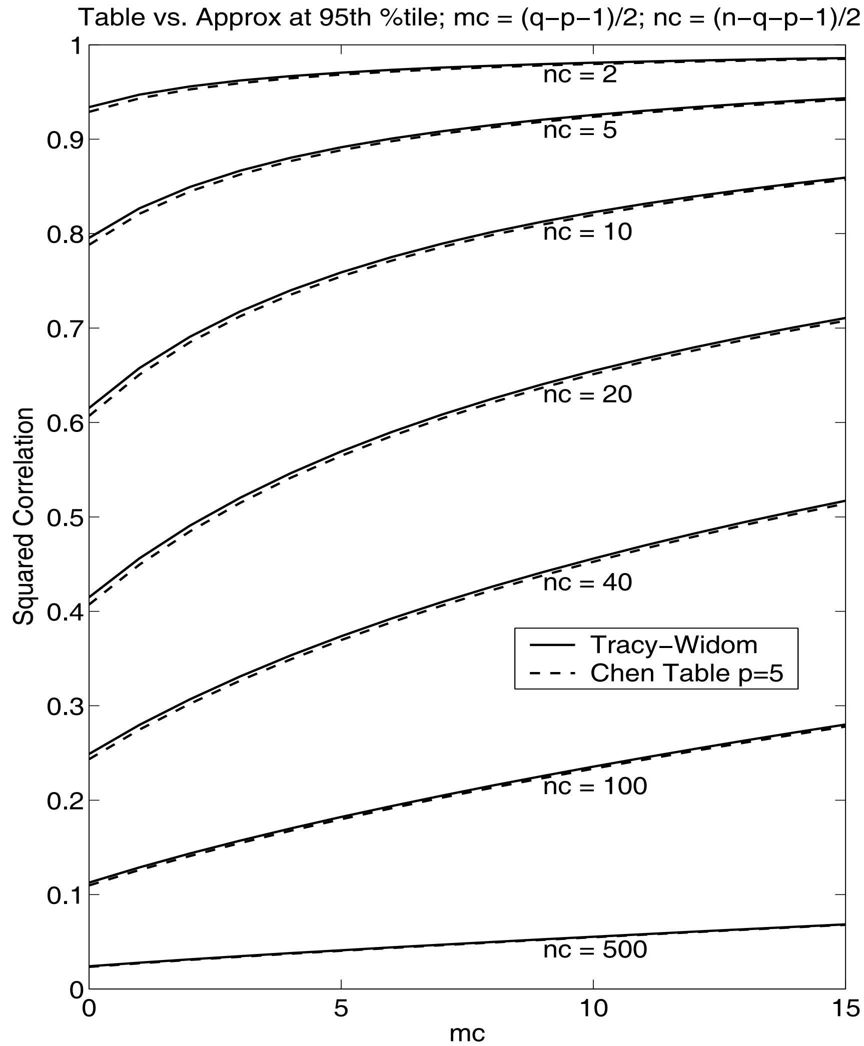
- corrections $(\mathbf{.5}, \mathbf{1}, \mathbf{2})$ improve approx'n for p, q small,
- $\Rightarrow$ error is $O(p^{-2/3})$ [instead of $O(p^{-1/3})$]

Approximation vs. Tables for $p = 5$

Tables: William Chen, IRS, (2002)

$$m_c = \frac{q-p-1}{2} \in [0, 15],$$

$$n_c = \frac{n-q-p-1}{2} \in [1, 1000]$$



Remarks

- $p^{-2/3}$ scale of variability for u_1
- 95th %tile $\doteq \mu_{p+} + \sigma_{p+}$, 99th %tile $\doteq \mu_{p+} + 2\sigma_{p+}$
- if $\mu_{p+} > .7$, **logit scale** $v_i = \log u_i/(1-u_i)$ better.

- **Smallest** eigenvalue: with previous assumptions and $\gamma_0 < \phi_0$, $\sigma_{p-}^3 = \frac{1}{(2n-2)^2} \frac{\sin^4(\phi_p - \gamma_p)}{\sin\phi_p \ \sin\gamma_p}$ then

$$\frac{\mu_{p-} - u_p}{\sigma_{p-}} \xrightarrow{\mathcal{D}} W_1 \qquad (W_2)$$

- Corresponding limit distributions for $u_2 \geq \cdots \geq u_k$, $u_{p-k} \geq \cdots \geq u_{p-1}$, k **fixed**

Agenda

- Classical multivariate techniques
- Hypothesis Testing: Single and Double Wishart
- Eigenvalue densities
- Linear Statistics
- Largest Eigenvalue
- Concluding Remarks

Concluding Remarks

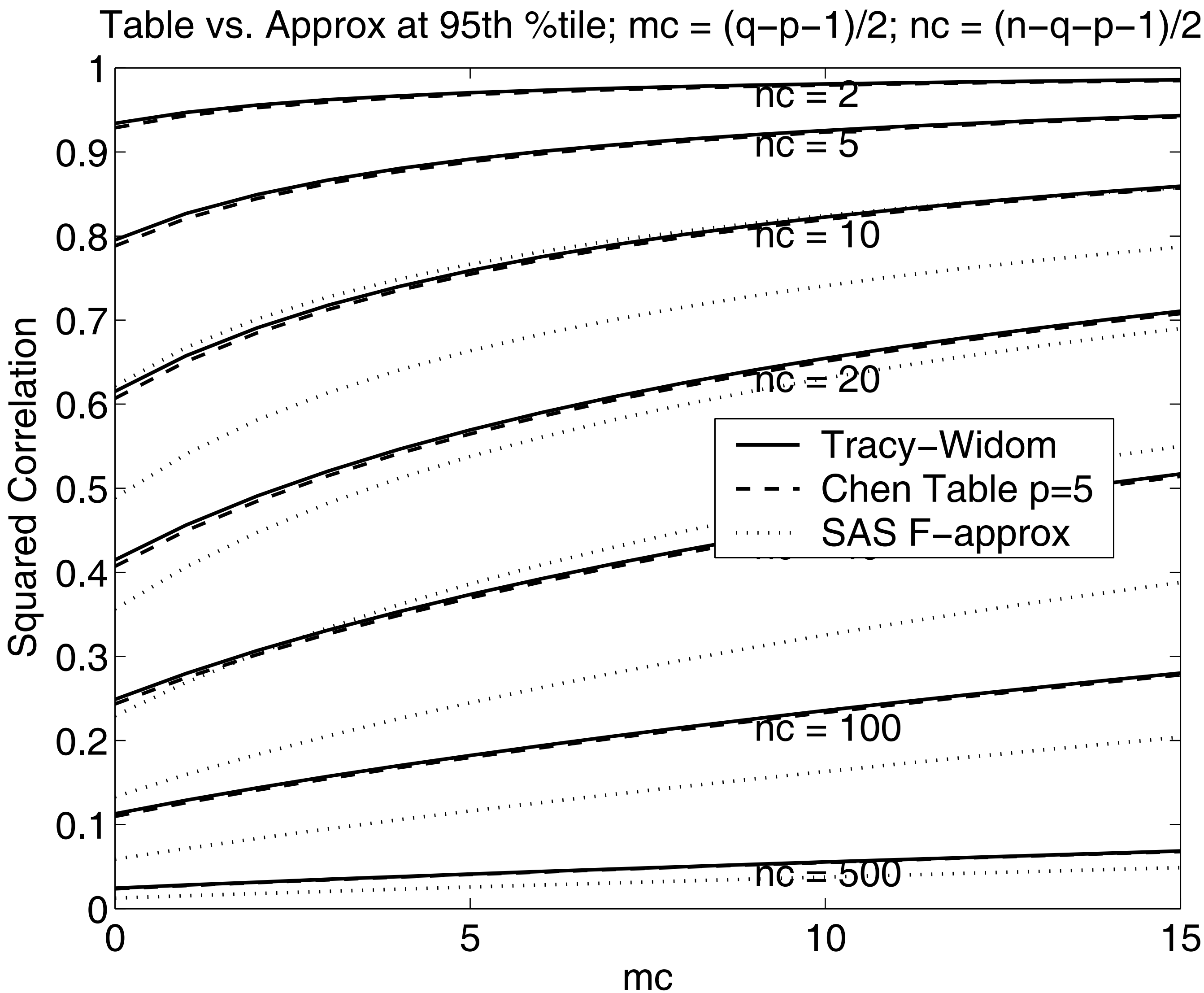
Numerous other topics deserve attention:

- distributions under alternative hypotheses: integral representations, matrix hypergeometric functions
- empirical distributions and graphical display (Wachter)
- computational advances: (Dumitriu, Edelman, Koev, Rao)
  - operations on random matrices
  - multivariate orthogonal polynomials
  - matrix hypergeometric functions
- estimation and testing for eigenvectors (Paul)
- technical role for RMT in other statistical areas: e.g. via large deviations results

Back-Up Slides

Upper Bound in SAS

Approximate $\frac{n-q}{q}\frac{u_1}{1-u_1}$ by $F_{q,n-q}$



Testing Subsequent Correlations

Suppose: $$\Sigma_{XY} = \begin{bmatrix} \rho_{\mathbf{1}}^{\mathbf{2}} & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & \rho_{\mathbf{p}}^{\mathbf{2}} & 0 & \cdots & 0 \end{bmatrix} \qquad p \le q, n-p$$

If largest r correlations are large, test

$$\mathbf{H_r} : \rho_{r+1} = \rho_{r+2} = \ldots = \rho_p = 0?$$

Comparison Lemma (from SVD interlacing)

$$\mathcal{L}(u_{r+1}|p, q, n; S_{XY} \in \mathbf{H_r}) \overset{st}{<} \mathcal{L}(u_1|p, q-r, n; \mathbf{I})$$

$\Rightarrow$ conservative $P-$values for H_r via

$$TW(p, q-r, n) \text{ approx'n to RHS}$$

[Aside: $\mathcal{L}(u_1|p-r, q-r, n; I)$ may be better, but no bounds]