Accurate and Efficient Traces of Beta–Wishart Matrices

Plamen Koev
San Jose State University

joint with Raymond Kan

NASCA 2018, Kalamata, Greece
July 4, 2018

What is a Wishart matrix?

- Importance comes from Multivariate Statistics
- Given: (n) observations on several (m) objects (thought to possess certain interdependence)

$$n \quad \overbrace{}^{m} \underbrace{\begin{bmatrix} * & * & \cdots & * \\ * & * & \cdots & * \\ * & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & * \end{bmatrix}}_{n \times m \text{ matrix of observations, } A}$$

- Questions:
  - Existence and nature of interdependence
  - Samples from same multivariate distribution?
  - Or can be classified?

Idealized (real, $\beta = 1$) model

$$M = \begin{bmatrix} N(0,1) & N(0,1) & \cdots & N(0,1) \\ N(0,1) & N(0,1) & \cdots & N(0,1) \\ N(0,1) & N(0,1) & \cdots & N(0,1) \\ N(0,1) & N(0,1) & \cdots & N(0,1) \\ \vdots & \vdots & & \vdots \\ N(0,1) & N(0,1) & \cdots & N(0,1) \end{bmatrix} \times \Sigma$$

where $\Sigma =$ covariance.

- $W \equiv M^T M$ is called *Wishart matrix*
- Tests: $\Sigma = I$? or, in general, $\Sigma = \Sigma_0$?
- Translated into questions on the p-value of $\lambda_{\max}$, trace, etc. of observed matrix $A^T A$ vs those of idealized Wishart model, W
- So we need densities of $\lambda_{\max}(W)$, $\underline{\text{trace}(W)}$, etc.

Complex Wishart $\beta = 2$

$$M = \begin{bmatrix} N(0,1) + i \cdot N(0,1) & \cdots & N(0,1) + i \cdot N(0,1) \\ N(0,1) + i \cdot N(0,1) & \cdots & N(0,1) + i \cdot N(0,1) \\ \vdots & \vdots & \vdots \\ N(0,1) + i \cdot N(0,1) & \cdots & N(0,1) + i \cdot N(0,1) \end{bmatrix} \times \Sigma$$

where $\Sigma =$ covariance.

- Quaternions, $\beta = 4$
- Models exist for any $\beta > 0$, called Beta-Wishart
- Division rings over the reals only for $\beta = 1, 2, 4$, but these are distributions, so OK
  - quaternions don't commute, but *random* ones do (joint density of *pq* and *qp* is the same!)
  - Another talk for another day

Densities of $\lambda_{max}, \lambda_{\min}$, trace of Wishart known but hard

- Today: Density of trace of Beta–Wishart (Muirhead, 1982):

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{\infty}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)}\sum_{\kappa\vdash k}\frac{(a)_\kappa}{k!}C_\kappa^{(\beta)}(I_m - z\Sigma^{-1})$$

- $a \equiv (n\beta)/2$; z – arbitrary
- infinite series ...
- $\kappa \vdash k$ means $\kappa = (\kappa_1, \ldots, \kappa_m)$ is an *integer partition* of k:
  $\kappa_1 + \cdots + \kappa_m = k$
  $\kappa_1 \geq \cdots \geq \kappa_m \geq 0$
- $C_\kappa^{(\beta)}(X) =$ Jack function, β generalization of Schur function (symmetric homogeneous polynomial in the eigs of X)
- $(a)_\kappa = \prod_{(i,j)\in\kappa}(a + i - 1 - \beta(j-1))$, Pochhammer symbol

Computational challenges

- Compute instead truncation (for large enough N)

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{N}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)k!}\sum_{\kappa\vdash k}(a)_{\kappa}C_{\kappa}^{(\beta)}(I_m - z\Sigma^{-1}),$$

- Key problems:
  - Number of partitions $|\kappa| \leq N$ is $\mathcal{O}(N^m)$, exponential ($m =$size of Wishart matrix)
  - Computing each Jack function “naively” also exponential

Partitions

- $1 = 1$
- $2 = 2$
  $2 = 1 + 1$
- $3 = 3$
  $3 = 2 + 1$
  $3 = 1 + 1 + 1$
- $4 = 4$
  $4 = 3 + 1$
  $4 = 2 + 2$
  $4 = 2 + 1 + 1$
  $4 = 1 + 1 + 1 + 1$
- $\ldots$
- In general, number of partitions of N in not more than m parts is $\mathcal{O}(N^m)$

Cost of Computing Jack Functions Naively ($\beta = 1$)

Degree	Partition κ	C_κ	Number of terms
1	(1)	$x_1 + \cdots + x_m$	$\mathcal{O}(m)$
2	(2)	$\sum_{i \leq j} x_i x_j$	$\mathcal{O}(m^2)$
3	$(1, 1, 1)$	$\sum_{i<j<k} x_i x_j x_k$	$\mathcal{O}(m^3)$
$\lvert\kappa\rvert$	κ	$\sum x_1^{\kappa_1} \cdots x_m^{\kappa_m}$	$\mathcal{O}\left(m^{\lvert\kappa\rvert}\right)$

New result: Density is sum over partitions of one part

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{N}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)k!}\sum_{\kappa\vdash k}(a)_{\kappa}C_{\kappa}^{(\beta)}(I_m - z\Sigma^{-1})$$

New result: Density is sum over partitions of one part

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{N}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)k!}\underbrace{\sum_{\kappa\vdash k}(a)_{\kappa}C_{\kappa}^{(\beta)}(I_m-z\Sigma^{-1})}_{(a)_k\cdot C_{(k)}^{(2a)}(I_m-z\Sigma^{-1})}$$

i.e.,

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{N}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)k!}(a)_k C_{(k)}^{(2a)}(I_m-z\Sigma^{-1})$$

- Results from combinatorial properties of Jack functions
- Still hard; for $\beta=1$ these are complete symmetric polys

Computing complete sym polynomials in $\mathcal{O}(m)$ time

- $C_{(n)}$ "contains" $C_{(k)}$ for $k < n$ Redundancy.

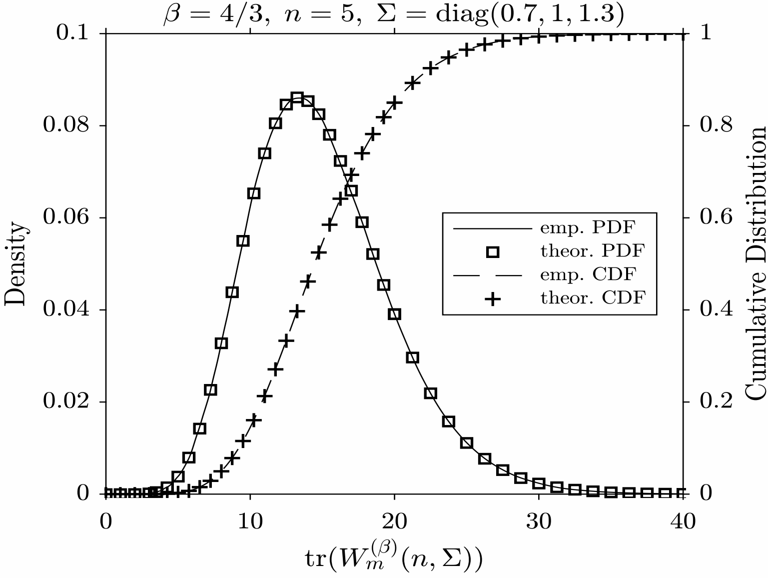
$$C_{(2)}(x_1, \ldots, x_m)$$

$$= \sum_{i \leq j} x_i x_j \qquad \left(\sim m^2 \text{ operations} \right)$$

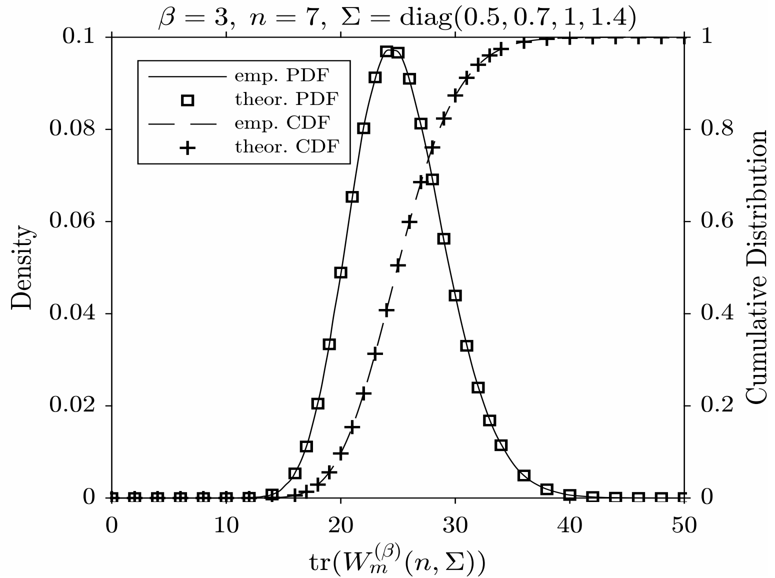
$$= \underbrace{x_1}_{s_1} x_1 + (\underbrace{x_1 + x_2}_{s_2}) x_2 + (\underbrace{x_1 + x_2 + x_3}_{s_3}) x_3 + \cdots + (\underbrace{x_1 + \cdots + x_m}_{s_m}) x_m$$

- New cost: $\mathcal{O}(m)$ instead of $\mathcal{O}(m^2)$
- Generalizes to all κ:
  Cost of $C_k(x_1, \ldots, x_m)$ now $\mathcal{O}(m)$ instead of $\mathcal{O}(m^k)$
- Overall cost of computing the trace $O(Nm)$, where
  N =degree of truncaton, m =size of matrix

Example 1, $\beta = 4/3$



Example 2, $\beta = 3$



Practical results: density of trace of Wishart

- 4×4 Wishart matrix, $N = 100$

	In 1 parts	In ≤ 4 parts
# partitions of ≤ 100	100	$239,279$
Actual time to compute	0.09 sec	1.14 sec

- 5×5 Wishart matrix, $N = 100$

	In 1 parts	In ≤ 5 parts
# partitions of ≤ 100	100	$1,209,293$
Actual time to compute	0.10 sec	25.12 sec

Why also "accurate" in floating point arithmetic?

- $\mathrm{fl}(a \odot b) = (a \odot b)(1+\delta), \quad \odot \in \{+,-,\times,/\}$
- Relative accuracy preserved in $\times, +, /$
  Proof: $(1+\delta)$ factors accumulate multiplicatively
- Subtractions of approximate quantities dangerous:

$$\begin{array}{rl} & .123456789\textit{xxx} \\ - & .123456789\textit{yyy} \\ \hline & .000000000\textit{zzz} \end{array}$$

- if *all* subtractions avoided, we get accuracy

Why also "accurate" in floating point arithmetic?

- In expression for density

$$\frac{e^{-\frac{x\beta}{2z}}}{x}\left|\frac{x\beta}{2}\Sigma^{-1}\right|^{a}\sum_{k=0}^{N}\frac{\left(\frac{x\beta}{2z}\right)^{k}}{\Gamma(ma+k)k!}(a)_{k}C_{(k)}^{(2a)}(I_m - z\Sigma^{-1})$$

  picking $z = \lambda_{\min}(\Sigma)$ makes it sum of positives
- Thus accurate

New results: Densities (red = new results)

WISHART

$\lambda_{\max}$ (cdf) $\quad \sim \left|\frac{x}{2}\beta\Sigma^{-1}\right|^{\frac{n}{2}\beta} e^{-\mathrm{tr}(\frac{x}{2}\beta\Sigma^{-1})} {}_1F_1^{(\beta)}(r; \frac{n}{2}\beta + r; \frac{x}{2}\beta\Sigma^{-1})$

$\lambda_{\min}$ (cdf) $\quad 1 - e^{\mathrm{tr}(-\frac{x}{2}\beta\Sigma^{-1})} \sum_{\kappa \subseteq (m^t)} \frac{1}{|\kappa|!} C_\kappa^{(\beta)}(\frac{x}{2}\beta\Sigma^{-1})$

trace (pdf) $\quad \left|\frac{x\beta}{2}\Sigma^{-1}\right|^a \frac{e^{-z}}{x\Gamma(ma)} {}_1F_1^{(2a)}(a; ma; zI_m - \frac{x\beta}{2}\Sigma^{-1})$

LAGUERRE

$\lambda_{\max/\min}$(cdf) $\quad$ same as Wishart with $\Sigma = I$ and $n = \frac{2a}{\beta}$

$\lambda_{\max}$ (pdf) $\quad \frac{am\beta\Gamma_m^{(\beta)}(r)e^{-\frac{mx}{2}\beta}}{2\Gamma_m^{(\beta)}(a+r)} (\frac{\beta x}{2})^{am-1} {}_1F_1^{(\beta)}(\frac{m}{2}\beta + 1; a + r; \frac{x\beta}{2}I_{m-1})$

$\lambda_{\min}$ (pdf) $\quad \frac{m}{2}\beta \frac{\Gamma_m^{(\beta)}(r)}{\Gamma_m^{(\beta)}(a)} \left(\frac{x}{2}\beta\right)^{tm} e^{-\frac{mx}{2}\beta} {}_2F_0^{(\beta)}(\frac{m}{2}\beta + 1, -t; -\frac{2}{\beta x}I_{m-1})$

JACOBI

$\lambda_{\max}$ (cdf) $\quad \sim x^{ma} \cdot {}_2F_1^{(\beta)}(a, r - b; a + r; xI)$

$\lambda_{\max}$ (pdf) $\quad \sim ma(1-x)^{b-r}x^{ma-1} {}_2F_1^{(\beta)}(a - \frac{\beta}{2}, r - b; a + r; xI_{m-1})$

$\lambda_{\min}$ (cdf) $\quad 1 - F_{b,a}(1-x)$, where $F_{a,b}(x)$ is the cdf of $\lambda_{\max}$

$\lambda_{\min}$ (pdf) $\quad f_{b,a}(1-x)$, where $f_{a,b}(x)$ is the density of $\lambda_{\max}$

MANOVA

λ_{max}(cdf) $\quad \left|x\Omega((1-x)I + x\Omega)^{-1}\right|^{\frac{p\beta}{2}} \sum_{k=0}^{nt} \sum_{\kappa \vdash k, \kappa_1 \le t} \frac{1}{k!} \left(\frac{p\beta}{2}\right)_\kappa^{(\beta)} C_\kappa^{(\beta)}((1-x)((1-x)I + x\Omega)^{-1})$

λ_{max}(pdf) $\quad \frac{|\Omega|^{\frac{n\beta}{2}} x^{\frac{nm\beta}{2}}}{|I + x\Omega^{-1}|^{\frac{n+p}{2}\beta}} {}_3F_2^{(\beta)}\left(\frac{n+p}{2}\beta, \frac{m\beta}{2} + 1, \frac{m-1}{2}\beta, \frac{n\beta}{2} + r, \frac{m\beta}{2}, x(\Omega + xI)^{-1}\right)$

Resources

- Macdonald's handwritten manuscripts:
  - Hypergeometric Functions I
  - Hypergeometric Functions II, q-analogues

  (available on ArXIV)
- Drensky, Edelman, Genoar, Kan, K., The Densities and Distributions of the Largest Eigenvalue and the Trace of a Beta-Wishart Matrix, 2018.
- Slides, papers, software: `math.sjsu.edu/~koev`