Wishart Distribution

Max Turgeon

STAT 7200–Multivariate Statistics

Objectives

- Understand the distribution of covariance matrices
- Understand the distribution of the MLEs for the multivariate normal distribution
- Understand the distribution of *functionals* of covariance matrices
- Visualize covariance matrices and their distribution
- Perform Box-Cox transformations to obtain multivariate normal data

Before we begin… i

- In this section, we will discuss *random matrices*
  - Therefore, we will talk about distributions, derivatives and integrals over *sets of matrices*
- It can be useful to identify the space $M_{n,p}(\mathbb{R})$ of $n \times p$ matrices with $\mathbb{R}^{np}$.
  - We can define the function $\mathrm{vec} : M_{n,p}(\mathbb{R}) \to \mathbb{R}^{np}$ that takes a matrix M and maps it to the np-dimensional vector given by concatenating the columns of M into a single vector.

$$\mathrm{vec}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = (1, 2, 3, 4).$$

Before we begin... ii

- Another important observation: structural constraints (e.g. symmetry, positive definiteness) reduce the number of "free" entries in a matrix and therefore the dimension of the subspace.
    - E.g. If A is a symmetric $p \times p$ matrix, there are only $\frac{1}{2}p(p+1)$ independent entries: the entries on the diagonal, and the off-diagonal entries above the diagonal (or below).

Wishart distribution i

- Let S be a random, positive semidefinite matrix of dimension $p \times p$.
  - We say S follows a *standard Wishart distribution* $W_p(m)$ if we can write
  $$S = \sum_{i=1}^{m} \mathbf{Z}_i \mathbf{Z}_i^T, \quad \mathbf{Z}_i \sim N_p(0, I_p).$$
  - We say S follows a *Wishart distribution* $W_p(m, \Sigma)$ with scale matrix Σ if we can write
  $$S = \sum_{i=1}^{m} \mathbf{Y}_i \mathbf{Y}_i^T, \quad \mathbf{Y}_i \sim N_p(0, \Sigma).$$

Wishart distribution ii

- We say S follows a *non-central Wishart distribution* $W_p(m, \Sigma; \Delta)$ with scale matrix Σ and non-centrality parameter Δ if we can write

$$S = \sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T, \quad \mathbf{Y}_i \sim N_p(\mu_i, \Sigma), \quad \Delta = \sum_{i=1}^m \mu_i \mu_i^T.$$

Example i

- Let $S \sim W_p(m)$ be Wishart distributed, with scale matrix $\Sigma = I_p$.
- We can therefore write $S = \sum_{i=1}^m \mathbf{Z}_i \mathbf{Z}_i^T$, with $\mathbf{Z}_i \sim N_p(0, I_p)$.

Example ii

- Using the properties of the trace, we have

$$
\begin{aligned}
\operatorname{tr}(S) &= \operatorname{tr}\left(\sum_{i=1}^{m} \mathbf{Z}_i \mathbf{Z}_i^T\right) \\
&= \sum_{i=1}^{m} \operatorname{tr}\left(\mathbf{Z}_i \mathbf{Z}_i^T\right) \\
&= \sum_{i=1}^{m} \operatorname{tr}\left(\mathbf{Z}_i^T \mathbf{Z}_i\right) \\
&= \sum_{i=1}^{m} \mathbf{Z}_i^T \mathbf{Z}_i.
\end{aligned}
$$

- Recall that $\mathbf{Z}_i^T \mathbf{Z}_i \sim \chi^2(p)$.

Example iii

- Therefore $\mathrm{tr}\,(S)$ is the sum of m independent copies of a $\chi^2(p)$, and so we have

$$\mathrm{tr}\,(S) \sim \chi^2(mp).$$

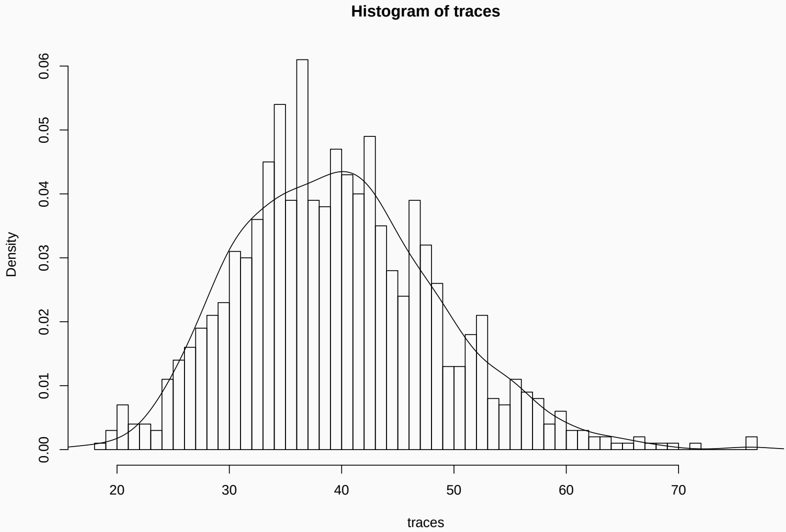
```
B <- 1000
n <- 10; p <- 4

traces <- replicate(B, {
  Z <- matrix(rnorm(n*p), ncol = p)
  W <- crossprod(Z)
  sum(diag(W))
})
```

Example iv

```r
hist(traces, 50, freq = FALSE)
lines(density(rchisq(B, df = n*p)))
```

Example v



Histogram of traces
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
20
30
40
50
60
70
traces

Non-singular Wishart distribution i

- As defined above, there is no guarantee that a Wishart variate is invertible.
- We will show that for $S \sim W_p(m, \Sigma)$ with Σ positive definite, S is invertible almost surely whenever $m \geq p$.

Lemma: Let Z be an $n \times n$ random matrix where the entries Z_{ij} are iid $N(0, 1)$. Then $P(\det Z = 0) = 0$.

Proof: We will prove this by induction on n. If $n = 1$, then the result hold since $N(0, 1)$ is absolutely continuous.

Now let $n > 1$ and assume the result holds for $n - 1$. Write

$$Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix},$$

where Z_{22} is $(n-1) \times (n-1)$. Note that by assumption, we have $\det Z_{22} \neq 0$ almost surely. Now, by the Schur determinant formula, we have

$$\begin{aligned} \det Z &= \det Z_{22} \det \left(Z_{11} - Z_{12} Z_{22}^{-1} Z_{21} \right) \\ &= (\det Z_{22}) \left(Z_{11} - Z_{12} Z_{22}^{-1} Z_{21} \right). \end{aligned}$$

Non-singular Wishart distribution iii

We now have

$$\begin{aligned}
P(|Z| = 0) &= P(|Z| = 0, |Z_{22}| \neq 0) + P(|Z| = 0, |Z_{22}| = 0) \\
&= P(|Z| = 0, |Z_{22}| \neq 0) \\
&= P(Z_{11} = Z_{12}Z_{22}^{-1}Z_{21}, |Z_{22}| \neq 0) \\
&= E\left(P(Z_{11} = Z_{12}Z_{22}^{-1}Z_{21}, |Z_{22}| \neq 0 \mid Z_{12}, Z_{22}, Z_{21})\right) \\
&= E(0) \\
&= 0.
\end{aligned}$$

Therefore, the result follows from induction. $\square$

Non-singular Wishart distribution iv

We are now ready to prove the main result: let $S \sim W_p(m, \Sigma)$ with $\det \Sigma \neq 0$, and write $S = \sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T$, with $\mathbf{Y}_i \sim N_p(0, \Sigma)$. If we let $\mathbb{Y}$ be the $m \times p$ matrix whose i-th row is $\mathbf{Y}_i$. Then

$$S = \sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T = \mathbb{Y}^T \mathbb{Y}.$$

Now note that

$$\mathrm{rank}(S) = \mathrm{rank}(\mathbb{Y}^T \mathbb{Y}) = \mathrm{rank}(\mathbb{Y}).$$

Non-singular Wishart distribution v

Furthermore, if we write $\Sigma = LL^T$ using the Cholesky decomposition, then we can write

$$\mathbb{Z} = \mathbb{Y}L^T,$$

where the rows $\mathbf{Z}_i$ of $\mathbb{Z}$ are $N_p(0, I_p)$ and $\mathrm{rank}(\mathbb{Z}) = \mathrm{rank}(\mathbb{Y})$.

Finally, we have

$$\begin{aligned} \mathrm{rank}(S) &= \mathrm{rank}(\mathbb{Z}) \\ &\geq \mathrm{rank}(\mathbf{Z}_1, \ldots, \mathbf{Z}_p) \\ &= p \quad \text{(a.s.)}, \end{aligned}$$

where the last equality follows from our Lemma. Since $\operatorname{rank}(S) = p$ almost surely, it is invertible almost surely. $\square$

Definition
If $S \sim W_p(m, \Sigma)$ with Σ positive definite and $m \geq p$, we say that S follows a *nonsingular* Wishart distribution. Otherwise, we say it follows a *singular* Wishart distribution.

Additional properties i

Let $S \sim W_p(m, \Sigma)$.

- We have $E(S) = m\Sigma$.
- If B is a $q \times p$ matrix, we have

$$BSB^T \sim W_p(m, B\Sigma B^T).$$

- If $T \sim W_p(n, \Sigma)$, then

$$S + T \sim W_p(m + n, \Sigma).$$

Additional properties ii

Now assume we can partition S and Σ as such:

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with S_{ii} and Σ_{ii} of dimension $p_i \times p_i$. We then have

- $S_{ii} \sim W_{p_i}(m, \Sigma_{ii})$
- If $\Sigma_{12} = 0$, then S_{11} and S_{22} are independent.