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

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

$$S = \sum_{i=1}^{m} \mathbf{Z}_i \mathbf{Z}_i^T, \quad \mathbf{Z}_i \sim N_p(0, I_p) \text{ indep..}$$

    - We say S follows a *Wishart distribution* $W_p(m, \Sigma)$ with scale matrix Σ if we can write

$$S = \sum_{i=1}^{m} \mathbf{Y}_i \mathbf{Y}_i^T, \quad \mathbf{Y}_i \sim N_p(0, \Sigma) \text{ indep..}$$

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

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

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

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

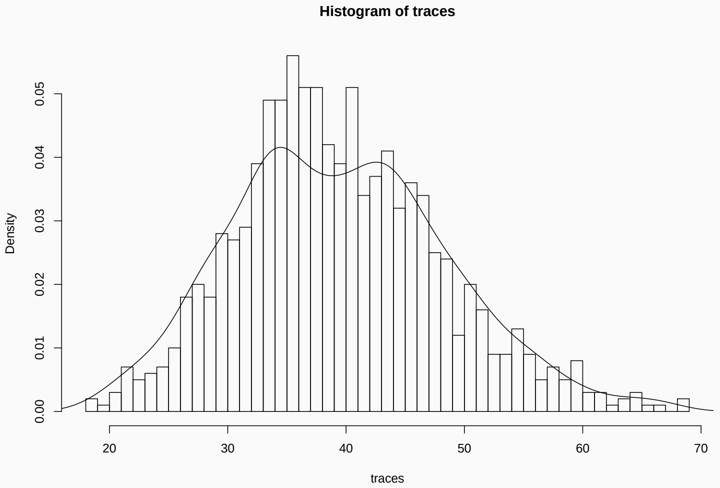
```
B <- 1000
n <- 10; p <- 4

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
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.
- **To show**: if $S \sim W_p(m, \Sigma)$ 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

Non-singular Wishart distribution ii

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

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

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

Non-singular Wishart distribution iv

where we used the laws of total probability (Line 1) and total expectation (Line 4). Therefore, the result follows from induction. $\square$

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

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

Non-singular Wishart distribution v

Now note that

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

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

$$\mathbb{Z} = \mathbb{Y}(L^{-1})^T,$$

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

Non-singular Wishart distribution vi

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

where the last equality follows from our Lemma. Since $\mathrm{rank}(S) = p$ 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.

Characteristic function i

- The definition of characteristic function can be extended to *random matrices*:
  - Let S be a $p \times p$ random matrix. The characteristic function of S evaluated at a $p \times p$ symmetric matrix T is defined as

$$\varphi_S(T) = E\left(\exp(i\mathrm{tr}(TS))\right).$$

- We will show that if $S \sim W_p(m, \Sigma)$, then

$$\varphi_S(T) = |I_p - 2i\Sigma T|^{-m/2}.$$

- First, we will use the Cholesky decomposition $\Sigma = LL^T$.

Characteristic function ii

- Next, we can write

$$S = L\left(\sum_{j=1}^{m}\mathbf{Z}_j\mathbf{Z}_j^T\right)L^T,$$

  where $\mathbf{Z}_j \sim N_p(0, I_p)$.
- Now, fix a symmetric matrix T. The matrix L^TTL is also symmetric, and therefore we can compute its spectral decomposition:

$$L^TTL = U\Lambda U^T,$$

  where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ is diagonal and $UU^T = I_p$.

Characteristic function iii

- We can now write

Characteristic function iv

$$
\begin{aligned}
\operatorname{tr}(TS) &= \operatorname{tr}\left(TL\left(\sum_{j=1}^{m}\mathbf{Z}_j\mathbf{Z}_j^T\right)L^T\right) \\
&= \operatorname{tr}\left(U\Lambda U^T\left(\sum_{j=1}^{m}\mathbf{Z}_j\mathbf{Z}_j^T\right)\right) \\
&= \operatorname{tr}\left(\Lambda U^T\left(\sum_{j=1}^{m}\mathbf{Z}_j\mathbf{Z}_j^T\right)U\right) \\
&= \operatorname{tr}\left(\Lambda\left(\sum_{j=1}^{m}(U^T\mathbf{Z}_j)(U^T\mathbf{Z}_j)^T\right)\right).
\end{aligned}
$$

Characteristic function v

- Two key observations:
  - $U^T\mathbf{Z}_j \sim N_p(0, I_p)$;
  - $\mathrm{tr}\left(\Lambda \mathbf{Z}_j \mathbf{Z}_j^T\right) = \sum_{k=1}^{p} \lambda_k Z_{jk}^2$.
- Putting all this together, we get

$$
\begin{aligned}
E\left(\exp(i\mathrm{tr}(TS))\right) &= E\left(\exp\left(i\sum_{j=1}^{m}\sum_{k=1}^{p}\lambda_k Z_{jk}^2\right)\right) \\
&= \prod_{j=1}^{m}\prod_{k=1}^{p} E\left(\exp\left(i\lambda_k Z_{jk}^2\right)\right).
\end{aligned}
$$

Characteristic function vi

- But $Z_{jk}^2 \sim \chi^2(1)$, and so we have

$$\varphi_S(T) = \prod_{j=1}^m \prod_{k=1}^p \varphi_{\chi^2(1)}(\lambda_k).$$

- Recall that $\varphi_{\chi^2(1)}(t) = (1 - 2it)^{-1/2}$, and therefore we have

$$\varphi_S(T) = \prod_{j=1}^m \prod_{k=1}^p (1 - 2i\lambda_k)^{-1/2}.$$

Characteristic function vii

- Since $\prod_{k=1}^{p}(1-2i\lambda_k)^{-1/2} = |I_p - 2i\Lambda|^{-1/2}$, we then have

$$\begin{aligned}
\varphi_S(T) &= \prod_{j=1}^{m} |I_p - 2i\Lambda|^{-1/2} \\
&= |I_p - 2i\Lambda|^{-m/2} \\
&= |I_p - 2iU\Lambda U^T|^{-m/2} \\
&= |I_p - 2iL^T T L|^{-m/2} \\
&= |I_p - 2i\Sigma T|^{-m/2}
\end{aligned}$$

$\square$

Density of Wishart distribution

- Let $S \sim W_p(m, \Sigma)$ with Σ positive definite and $m \geq p$. The density of S is given by

$$f(S) = \frac{1}{2^{pm/2}\Gamma_p(\frac{m}{2})|\Sigma|^{m/2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)\right) |S|^{(m-p-1)/2},$$

where

$$\Gamma_p(u) = \pi^{p(p-1)/4} \prod_{i=0}^{p-1} \Gamma\left(u - \frac{i}{2}\right), \quad u > \frac{1}{2}(p-1).$$

- *Proof*: Compute the characteristic function using the expression for the density and check that we obtain the same result as before (**Exercise**).

Sampling distribution of sample covariance

- We are now ready to prove the results we stated a few lectures ago.
- Recall again the univariate case:
  - $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$;
  - $\bar{X}$ and s^2 are independent.
- In the multivariate case, we want to prove:
  - $(n-1)S_n \sim W_p(n-1, \Sigma)$;
  - $\bar{\mathbf{Y}}$ and S_n are independent.
- We will show that using the **multivariate Cochran theorem**

Cochran theorem

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample with $\mathbf{Y}_i \sim N_p(0, \Sigma)$, and write $\mathbb{Y}$ for the $n \times p$ matrix whose i-th row is $\mathbf{Y}_i$. Let A, B be $n \times n$ symmetric matrices, and let C be a $q \times n$ matrix of rank q. Then

1. $\mathbb{Y}^T A \mathbb{Y} \sim W_p(m, \Sigma)$ if and only if $A^2 = A$ and $\mathrm{tr} A = m$.
2. $\mathbb{Y}^T A \mathbb{Y}$ and $\mathbb{Y}^T B \mathbb{Y}$ are independent if and only if $AB = 0$.
3. $\mathbb{Y}^T A \mathbb{Y}$ and $C \mathbb{Y}$ are independent if and only if $CA = 0$.

Application i

- Let $C = \frac{1}{n}\mathbf{1}^T$, where $\mathbf{1}$ is the n-dimensional vector of ones.
- Let $A = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.
- Then we have

$$\mathbb{Y}^T A \mathbb{Y} = (n-1)S_n, \qquad C\mathbb{Y} = \bar{\mathbf{Y}}^T.$$

- We need to check the conditions of Cochran's theorem:
  - $A^2 = A$;
  - $CA = 0$;
  - $\mathrm{tr} A = n - 1$.

Application ii

- Using Parts 1. and 3. of the theorem, we can conclude that
  - $(n-1)S_n \sim W_p(n-1, \Sigma)$;
  - $\bar{\mathbf{Y}}$ and S_n are independent.

Proof (Cochran theorem) i

Note 1: We typically only use one direction ($\Leftarrow$).

Note 2: We will only prove the first part.

- Since A is symmetric, we can compute its spectral decomposition as usual:

$$A = U\Lambda U^T.$$

- By assuming $A^2 = A$, we are forcing the same condition on the eigenvalues:

$$\Lambda^2 = \Lambda.$$

- But only two real numbers are possible $\lambda_i \in \{0, 1\}$.

Proof (Cochran theorem) ii

- Given that $\mathrm{tr} A = m$, and after perhaps reordering the eigenvalues, we have

$$\lambda_1 = \cdots = \lambda_m = 1, \quad \lambda_{m-1} = \cdots = \lambda_n = 0.$$

- Now, set $\mathbb{Z} = U^T \mathbb{Y}$, and let $\mathbf{Z}_i$ be the i-th row of $\mathbb{Z}$. We have

$$\begin{aligned}
\mathrm{Cov}(\mathbb{Z}) &= E((U^T \mathbb{Y})^T (U^T \mathbb{Y})) \\
&= E(\mathbb{Y}^T U U^T \mathbb{Y}) \\
&= E(\mathbb{Y}^T \mathbb{Y}) \\
&= \mathrm{Cov}(\mathbb{Y}).
\end{aligned}$$

Proof (Cochran theorem) iii

- Therefore, the covariance structures of $\mathbb{Y}$ and $\mathbb{Z}$ are the same:
  - The vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are still independent.
  - $\mathbf{Z}_i \sim N_p(0, \Sigma)$.
- We can now write

$$\begin{aligned}\mathbb{Y}^T A \mathbb{Y} &= \mathbb{Y}^T U \Lambda U^T \mathbb{Y} \\ &= \mathbb{Z}^T \Lambda \mathbb{Z} \\ &= \sum_{i=1}^{m} \mathbf{Z}_i \mathbf{Z}_i^T.\end{aligned}$$

- Therefore, we conlude that $\mathbb{Y}^T A \mathbb{Y} \sim W_p(m, \Sigma)$. $\square$

Bartlett decomposition i

- Recall that the Wishart distribution is a distribution on the set of *positive semi-definite matrices*.
  - This implies symmetry and a non-negative eigenvalues.
- These constraints are natural for covariance matrices, but it forces dependence between the entries that can make computations difficult.
- The **Bartlett decomposition** is a reparametrization of the Wishart distribution in terms of $p(p+1)/2$ *independent* entries.
  - You can think of it as a *stochastic* version of the Cholesky decomposition.

Bartlett decomposition ii

- Let $S \sim W_p(m, \Sigma)$, where $m \geq p$ and Σ is positive definite, and write $S = LL^T$ using the Cholesky decomposition. Then the density of L is given by

$$f(L) = \frac{2^p}{K} \prod_{i=1}^{p} \ell_{ii}^{m-i} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}LL^T)\right),$$

  where $K = 2^{mp/2}|\Sigma|\Gamma_p(m/2)$ and ℓ_{ij} is the (i, j)-th entry of L.

Proof i

- This result will follow from the formula for the density after a transformation.
- Recall that the density of S is:

$$f(S) = \frac{1}{K}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)\right)|S|^{(m-p-1)/2}.$$

- Note that we have

$$\mathrm{tr}(\Sigma^{-1}S) = \mathrm{tr}(\Sigma^{-1}LL^T),$$
$$|S| = |LL^T| = |L|^2 = \prod_{i=1}^{p}\ell_{ii}^2.$$

Proof ii

- Putting all this together, we have

$$\begin{aligned}
f(LL^T) &= \frac{1}{K}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)\right)|S|^{(m-p-1)/2} \\
&= \frac{1}{K}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}LL^T)\right)\prod_{i=1}^{p}\ell_{ii}^{m-p-1}.
\end{aligned}$$

- To get the density of L, we need to multiply by the Jacobian of the inverse transformation $L \mapsto LL^T$.

Proof iii

- A simple yet tedious computation (see for example Theorem 2.1.9 in Muirhead) gives:

$$|J| = 2^p \prod_{i=1}^{p} \ell_{ii}^{p-i+1}.$$

- Finally, we get the expression we wanted:

$$\begin{aligned} f(L) &= \frac{2^p \prod_{i=1}^{p} \ell_{ii}^{p-i+1}}{K} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}LL^T)\right) \prod_{i=1}^{p} \ell_{ii}^{m-p-1} \\ &= \frac{2^p}{K} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}LL^T)\right) \prod_{i=1}^{p} \ell_{ii}^{m-i}. \end{aligned}$$

$\square$

Corollary i

If $\Sigma = I_p$, the elements ℓ_{ij} are all independent, and they follow the following distributions:

$$\ell_{ii}^2 \sim \chi^2(m - i + 1),$$
$$\ell_{ij} \sim N(0, 1), \quad i > j.$$

Proof:

- When $\Sigma = I_p$, the expression for $\mathrm{tr}(\Sigma^{-1}LL^T)$ simplifies:

$$\mathrm{tr}(\Sigma^{-1}LL^T) = \mathrm{tr}(LL^T) = \sum_{i \geq j} \ell_{ij}^2.$$

Corollary ii

- This allows us to rewrite the density $f(L)$ (up to a constant):

$$\begin{aligned} f(L) &\propto \exp\left(-\frac{1}{2}\mathrm{tr}(LL^T)\right)\prod_{i=1}^{p}\ell_{ii}^{m-i} \\ &= \exp\left(-\frac{1}{2}\sum_{i\geq j}\ell_{ij}^2\right)\prod_{i=1}^{p}\ell_{ii}^{m-i} \\ &= \left\{\prod_{i>j}\exp\left(-\frac{1}{2}\ell_{ij}^2\right)\right\}\left\{\prod_{i=1}^{p}\exp\left(-\frac{1}{2}\ell_{ii}^2\right)\ell_{ii}^{m-i}\right\}, \end{aligned}$$

Corollary iii

which is the product of the marginals we wanted. $\square$

Example i

```r
B <- 1000
n <- 10
p <- 5

bartlett <- replicate(B, {
  X <- matrix(rnorm(n*p), ncol = p)
  L <- chol(crossprod(X))
})

dim(bartlett)
```

Example ii

```
## [1]    5    5 1000
```

```r
library(tidyverse)

# Extract and plot diagonal^2
diagonal <- purrr::map_df(seq_len(B), function(i) {
  tmp <- diag(bartlett[,,i])^2
  data.frame(matrix(tmp, nrow = 1))
})
```

Example iii

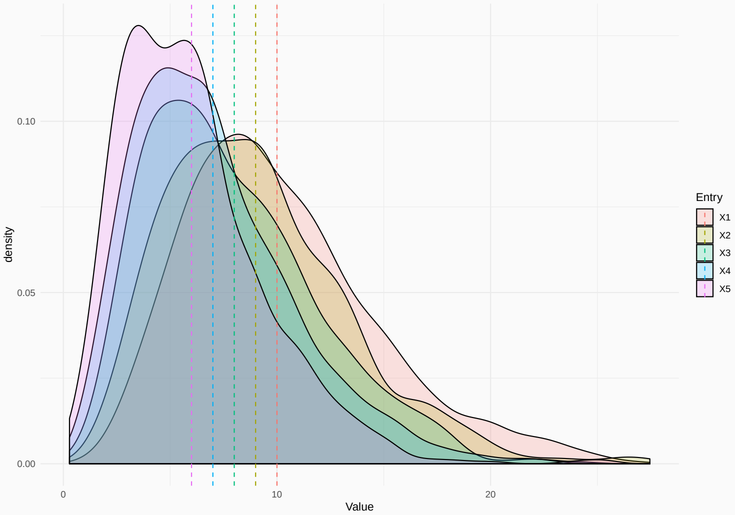
```r
# Put into long format
diag_plot <- gather(diagonal, Entry, Value)

# Add chi-square means
diag_means <- data.frame(
  Entry = paste0("X", seq_len(p)),
  mean = n - seq_len(p) + 1
)
```

Example iv

```r
ggplot(diag_plot, aes(Value, fill = Entry)) +
  geom_density(alpha = 0.2) +
  theme_minimal() +
  geom_vline(data = diag_means,
             aes(xintercept = mean,
                 colour = Entry),
             linetype = 'dashed')
```

Example v



Example vi

```r
# Extract and plot off-diagonal
off_diagonal <- purrr::map_df(seq_len(B), function(i)
  tmp <- bartlett[,,i][upper.tri(bartlett[,,i])]
  data.frame(matrix(tmp, nrow = 1))
})
dim(off_diagonal)
```

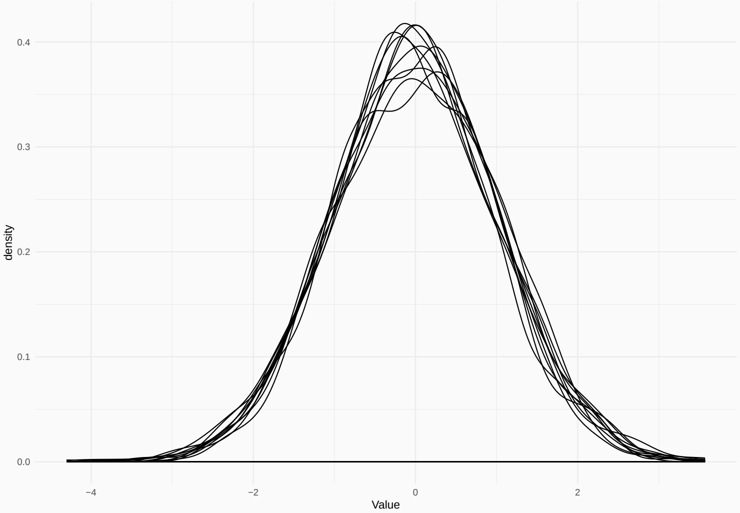
```
## [1] 1000   10
```

Example vii

```r
# Put into long format
offdiag_plot <- gather(off_diagonal, Entry, Value)
```

```r
ggplot(offdiag_plot, aes(Value, group = Entry)) +
  geom_density(fill = NA) +
  theme_minimal()
```

Example viii



Distribution of the Generalized Variance i

- As an application of the Bartlett decomposition, we will look at the distribution of the *generalized variance*:

$$GV(S) = |S|, \quad S \sim W_p(m, \Sigma).$$

- **Theorem**: If $S \sim W_p(m, \Sigma)$ with $m \geq p$ and Σ positive definite, then the ratio

$$GV(S)/GV(\Sigma) = |S|/|\Sigma|$$

follows the same distribution as a product of chi-square distributions:

$$\prod_{i=1}^{p} \chi^2(m - i + 1).$$

Distribution of the Generalized Variance ii

Proof:

- First, we have

$$\frac{|S|}{|\Sigma|} = |S||\Sigma^{-1}| = |\Sigma^{-1/2}||S||\Sigma^{-1/2}| = |\Sigma^{-1/2} S \Sigma^{-1/2}|.$$

- Moreover, we have that $\Sigma^{-1/2} S \Sigma^{-1/2} \sim W_p(m, I_p)$, so we can use the result of the Corollary above.
- If we write $\Sigma^{-1/2} S \Sigma^{-1/2} = LL^T$ using the Bartlett decomposition, we have

$$\frac{|S|}{|\Sigma|} = |LL^T| = |L|^2 = \prod_{i=1}^{p} \ell_{ii}^2.$$

Distribution of the Generalized Variance iii

- Our result follows from the characterisation of the distribution of ℓ_{ii}^2. $\square$
- **Note**: The distribution of $GV(S)/GV(\Sigma)$ does not depend on Σ.
  - It is a pivotal quantity.
- **Note 2**: If S_n is the sample covariance, then $(n-1)S_n \sim W_p(n-1, \Sigma)$ and therefore

$$(n-1)^p \frac{GV(S_n)}{GV(\Sigma)} \sim \prod_{i=1}^{p} \chi^2(n-i).$$

Example i

- We will use the Ramus dataset (see slides on *Multivariate normal*).
- We will construct a 95% confidence interval for the population generalized variance.
    - Under a multivariate normality assumption, which probably doesn't hold...

Example ii

```r
var_names <- c("Age8", "Age8.5",
               "Age9", "Age9.5")

dataset <- ramus[,var_names]
dim(dataset)
```

```
## [1] 20  4
```

Example iii

```r
# Sample covariance
Sn <- cov(dataset)

# Generalized variance
det(Sn)
```

```
## [1] 1.068328
```

Example iv

```r
# Simulate quantiles
set.seed(7200)
n <- nrow(dataset)
p <- ncol(dataset)
B <- 1000

simulated_vals <- replicate(B, {
  prod(rchisq(p, df = n - seq_len(p)))/((n-1)^p)
})
```

Example v

```r
bounds <- quantile(simulated_vals,
                   probs = c(0.025, 0.975))

bounds
```

```
##      2.5%     97.5%
## 0.1409302 2.0241338
```

Example vi

```
# 95% Confidence interval (reverse bounds)
det(Sn)/rev(bounds)
```

```
##     97.5%     2.5%
## 0.527795 7.580545
```

Visualization i

- Visualizing covariance/correlation matrices can be difficult, especially when the number of variables p increases.
  - One possibility is a **heatmap**, that assign a colour to the individual coariances/correlations.
- Visualizing *distributions* of random matrices is even harder
  - Already when $p = 2$, this is a 3-dimensional object...

Visualization ii

- One possibility is to decompose the distribution of a random matrix (or a sample thereof) into a series of univariate and bivariate graphical summaries. For example:
    - Histograms of the covariances/correlations;
    - Scatter plots for pairs of covariances;
    - Histograms of traces and determinants.

Example i

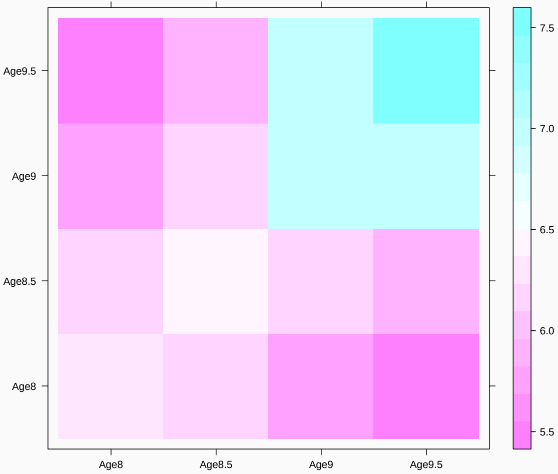
```r
# Recall our covariance matrix for the Ramus dataset
round(Sn, 2)
```

```
##         Age8 Age8.5 Age9 Age9.5
## Age8    6.33   6.19 5.78   5.55
## Age8.5 6.19   6.45 6.15   5.92
## Age9    5.78   6.15 6.92   6.95
## Age9.5 5.55   5.92 6.95   7.46
```

Example ii

```r
# Visually we get
lattice::levelplot(Sn, xlab = "", ylab = "")
```

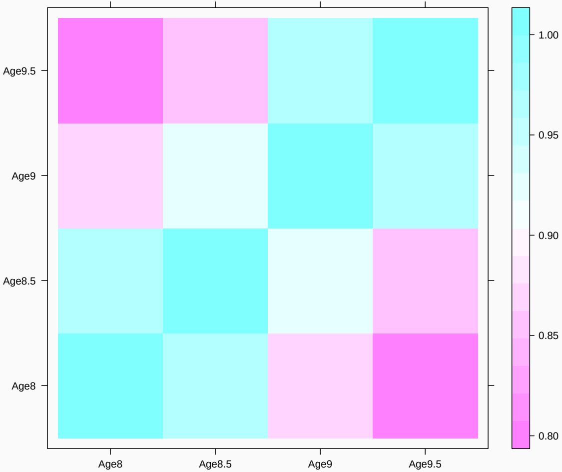
Example iii



Example iv

```r
# Perhaps easier to interpret as correlations
# But be careful with the scale!
lattice::levelplot(cov2cor(Sn),
                   xlab = "", ylab = "")
```

Example v



Age9.5
Age9
Age8.5
Age8
Age8
Age8.5
Age9
Age9.5
1.00
0.95
0.90
0.85
0.80

Example vi

Next, we will visualize the distribution of S_n using bootstrap.

```
B <- 1000
n <- nrow(dataset)

boot_covs <- lapply(seq_len(B), function(b) {
  data_boot <- dataset[sample(n, n, replace = TRUE),]
  return(cov(data_boot))
})
```

Example vii

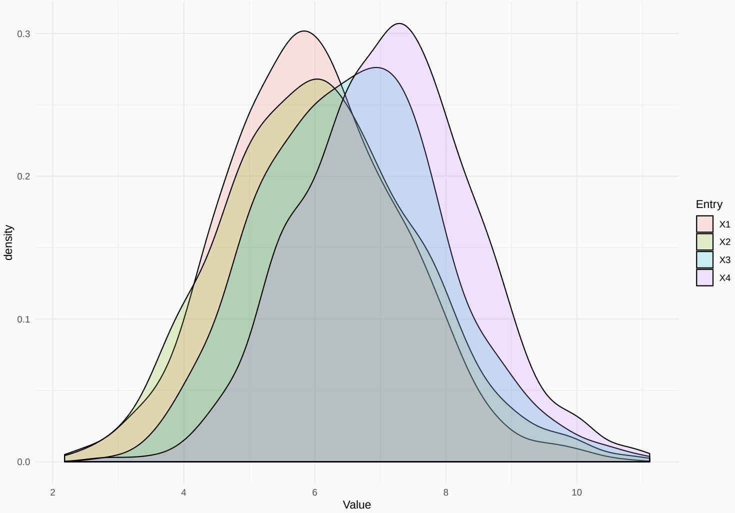
```r
# Extract the diagonal entries
diagonal <- purrr::map_df(boot_covs, function(Sn) {
  tmp <- diag(Sn)
  data.frame(matrix(tmp, nrow = 1))
  })
```

Example viii

```r
# Put into long format
diag_plot <- gather(diagonal, Entry, Value)

ggplot(diag_plot, aes(Value, fill = Entry)) +
  geom_density(alpha = 0.2) +
  theme_minimal()
```

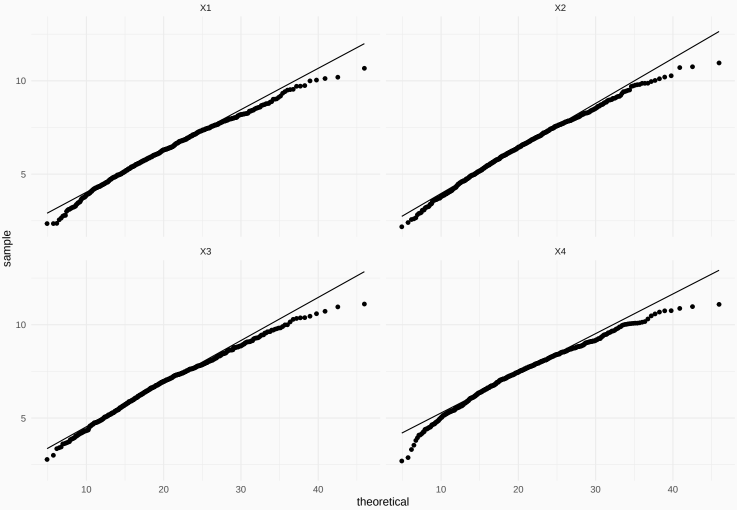
Example ix



Example x

```r
# Multivariate normal theory predicts
# the diagonal entry should be scaled chi-square
ggplot(diag_plot, aes(sample = Value)) +
  geom_qq(distribution = qchisq,
          dparams = list(df = n - 1)) +
  theme_minimal() + facet_wrap(~ Entry) +
  geom_qq_line(distribution = qchisq,
               dparams = list(df = n - 1))
```

Example xi



Example xii

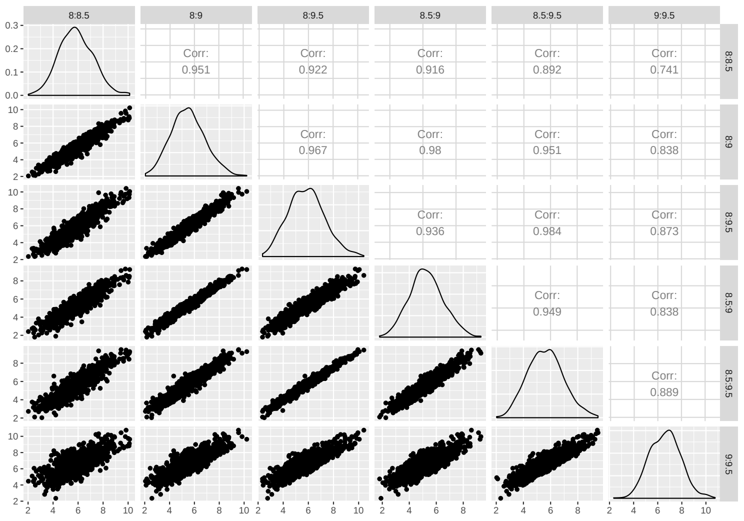
```r
# Finally, let's look at pairwise scatterplots
# for off-diagonal entries
off_diag <- purrr::map_df(boot_covs, function(Sn) {
  tmp <- Sn[upper.tri(Sn)]
  data.frame(matrix(tmp, nrow = 1))
  })
```

Example xiii

```r
# Add column names
names(off_diag) <- c(paste0("8:",c("8.5","9","9.5")),
                     paste0("8.5:",c("9","9.5")),
                     "9:9.5")

GGally::ggpairs(off_diag)
```

Example xiv



8:8.5
8:9
8:9.5
8.5:9
8.5:9.5
9:9.5
Corr: 0.951
Corr: 0.922
Corr: 0.916
Corr: 0.892
Corr: 0.741
Corr: 0.967
Corr: 0.98
Corr: 0.951
Corr: 0.838
Corr: 0.936
Corr: 0.984
Corr: 0.873
Corr: 0.949
Corr: 0.838
Corr: 0.889

Summary

- **Wishart random matrices** are sums of outer products of independent multivariate normal variables with the same scale matrix Σ.
- They allow us to give a description of the sample covariance matrices and its *functionals*:
  - E.g. trace, generalized variance, etc.
- The **Bartlett decomposition** gives us a reparametrization of the Wishart distribution with independent constaints of the entries.
  - Positive diagonal entries; contant zero above the diagonal; unconstrained below the diagonal.