STAT 730 Chapter 3: Normal Distribution Theory

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 730: Multivariate Analysis

Nice properties of multivariate normal random vectors

- Multivariate normal easily generalizes univariate normal. Much harder to generalize Poisson, gamma, exponential, etc.
- Defined completely by first and second moments, i.e. mean vector and covariance matrix.
- If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\sigma_{ij} = 0$ implies x_i independent of x_j.
- $\mathbf{a}'\mathbf{x} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$.
- Central Limit Theorem says sample means are approximately multivariate normal.
- Simple geometry makes properties intuitive.

Definition via Cramér-Wold

$\mathbf{x}$ *is multivariate normal* $\Leftrightarrow$ $\mathbf{a}'\mathbf{x}$ *is normal for all* $\mathbf{a}$.

<u>def'n</u> $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{a}'\mathbf{x} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for all $\mathbf{a} \in \mathbb{R}^p$.

<u>thm</u>: If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then its characteristic function is $\phi_{\mathbf{x}}(\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$.

<u>Proof</u>: Let $y = \mathbf{t}'\mathbf{x}$. Then the c.f. of y is

$$\phi_y(s) \stackrel{def}{=} E\{e^{isy}\} = \exp\{isE(y) - \tfrac{1}{2}s^2 var(y)\} = \exp\{is\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}s^2\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}.$$

Then the c.f. of $\mathbf{x}$ is

$$\phi_{\mathbf{x}}(\mathbf{t}) \stackrel{def}{=} E\{e^{i\mathbf{t}'\mathbf{x}}\} = \phi_y(1) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}).\square$$

Using the c.f. we see that if $\boldsymbol{\Sigma} = \mathbf{0}$ then $\mathbf{x} = \boldsymbol{\mu}$ with probability one, i.e. $N_p(\boldsymbol{\mu}, \mathbf{0}) = \delta_{\boldsymbol{\mu}}$.

Linear transformations of $\mathbf{x}$ are also normal

thm: $\mathbf{x} \sim N_p(\mathbf{x}, \mathbf{\Sigma})$, $\mathbf{A} \in \mathbb{R}^{q \times p}$, and $\mathbf{c} \in \mathbb{R}^q$
$\Rightarrow \mathbf{Ax} + \mathbf{c} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A\Sigma A}')$.

Proof: Let $\mathbf{b} \in \mathbb{R}^q$; then $\mathbf{b}'[\mathbf{Ax} + \mathbf{c}] = [\mathbf{b}'\mathbf{A}]\mathbf{x} + \mathbf{b}'\mathbf{c}$. Since $[\mathbf{b}'\mathbf{A}]\mathbf{x}$ is univariate normal by def'n, $[\mathbf{b}'\mathbf{A}]\mathbf{x} + \mathbf{b}'\mathbf{c}$ is also for any $\mathbf{b}$. The specific forms for the mean and covariance are standard results for any $\mathbf{Ax} + \mathbf{c}$ (Chapter 2). $\square$

Corollary: Any subset of $\mathbf{x}$ is multivariate normal; the x_i are normal.

Note: you will show $\phi_y(t) = e^{it\boldsymbol{\mu} - \sigma^2 t^2/2}$ for $y \sim N(\mu, \sigma^2)$ in your HW.

Normality and independence

Let $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$ of dimension k and $p - k$. Also partition $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$. Then $\mathbf{x}_1$ indep. $\mathbf{x}_2 \Leftrightarrow C(\mathbf{x}_1, \mathbf{x}_2) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = \mathbf{0}$.

Proof:

$$\begin{aligned} \phi_{\mathbf{x}}(\mathbf{t}) &= \phi_{\mathbf{x}_1}(\mathbf{t}_1)\phi_{\mathbf{x}_2}(\mathbf{t}_2) = \exp(i\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2 - \tfrac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 - \tfrac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2) \\ &\Leftrightarrow C(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{0}. \square \end{aligned}$$

Some results based on last two slides

Corollary: $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \mathbf{y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathcal{I}_n)$ and $U = (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}'\mathbf{y} \sim \chi^2_p$.

Corollary: $\mathbf{x} \sim N_p(\mathbf{0}, \mathcal{I}) \Rightarrow \frac{\mathbf{a}'\mathbf{x}}{||\mathbf{a}||} \sim N(0, 1)$ for $\mathbf{a} \neq \mathbf{0}$.

thm: Let $\mathbf{A} \in \mathbb{R}^{n_1 \times p}$, $\mathbf{B} \in \mathbb{R}^{n_2 \times p}$, and $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\mathbf{Ax}$ indep. $\mathbf{Bx} \Leftrightarrow \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$.

Last one is immediate from previous two slides by finding the distribution of $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{x}$.

Corollary: $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \sigma^2\mathcal{I})$ and $\mathbf{G}\mathbf{G}' = \mathcal{I}$ then $\mathbf{Gx} \sim N_p(\mathbf{G}\boldsymbol{\mu}, \sigma^2\mathcal{I})$. Also $\mathbf{Gx}$ indep. of $(\mathcal{I} - \mathbf{G}'\mathbf{G})\mathbf{x}$.

Conditional distribution of $\mathbf{x}_2|\mathbf{x}_1$

Let $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$ of dimension k and $p-k$. Also partition $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$. Let $\mathbf{x}_{2.1} = \mathbf{x}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1$.

$$\begin{aligned}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_{2.1} \end{bmatrix} &= \begin{bmatrix} \mathcal{I}_k & \mathbf{0} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathcal{I}_{p-k} \end{bmatrix} \mathbf{x} \\ &\sim N_p\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{bmatrix} \right).\end{aligned}$$

So $\mathbf{x}_1$ indep. $\mathbf{x}_{2.1}$. Then $\mathbf{x}_2|\mathbf{x}_1 = \mathbf{x}_{2.1} + \underbrace{\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1}_{\text{constant}}$ has distribution...

thm: $\mathbf{x}_2|\mathbf{x}_1 \sim N_{p-k}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$.

Very useful! Mean and variance results hold for non-normal $\mathbf{x}$ too.

Transformations of normal data matrix

If $\mathbf{x}_1, \ldots, \mathbf{x}_n \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]'$ is a $n \times p$ "normal data matrix."

General transformations are of the form **AXB**. An important example is $\bar{\mathbf{x}}' = [\frac{1}{n}\mathbf{1}_n']\mathbf{X}[\mathcal{I}]$, the sample mean. One can show via c.f. that...

<u>thm</u>: $\mathbf{x}_1, \ldots, \mathbf{x}_n \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.

thm: If $\mathbf{X}(n \times p)$ is data matrix from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y}(m \times q) = \mathbf{AXB}$ then $\mathbf{Y}$ is normal data matrix $\Leftrightarrow$

(a) $\mathbf{A1}_n = \alpha \mathbf{1}_m$ for $\alpha \in \mathbb{R}$, or $\mathbf{B}'\boldsymbol{\mu} = \mathbf{0}$, and

(b) $\mathbf{AA}' = \beta \boldsymbol{\mathcal{I}}_p$ some $\beta \in \mathbb{R}$, or $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$.

We will prove this in class. Some necessary results follow.

def'n: For any matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, let

$$\mathbf{X}^v = \begin{bmatrix} \mathbf{x}_{(1)} \\ \vdots \\ \mathbf{x}_{(p)} \end{bmatrix} = (\mathbf{x}'_{(1)}, \ldots, \mathbf{x}'_{(p)})' \in \mathbb{R}^{np}.$$

Kronecker products

<u>def'n</u> Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{p \times q}$. Then

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{np \times mq}.$$

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $C(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij}\boldsymbol{\Sigma}$, so

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \sim N_{np}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} \right) = N_{np}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathcal{I}_n \otimes \boldsymbol{\Sigma}).$$

Kronecker products, dist'n of $\mathbf{X}^v$

prop: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\begin{aligned} \mathbf{X}^v &= \begin{bmatrix} \mathbf{x}_{(1)} \\ \mathbf{x}_{(2)} \\ \vdots \\ \mathbf{x}_{(p)} \end{bmatrix} \sim N_{np}\left(\begin{bmatrix} \mu_1 \mathbf{1}_n \\ \mu_2 \mathbf{1}_n \\ \vdots \\ \mu_p \mathbf{1}_n \end{bmatrix}, \begin{bmatrix} \sigma_{11}\boldsymbol{\mathcal{I}}_n & \sigma_{12}\boldsymbol{\mathcal{I}}_n & \cdots & \sigma_{1p}\boldsymbol{\mathcal{I}}_n \\ \sigma_{21}\boldsymbol{\mathcal{I}}_n & \sigma_{22}\boldsymbol{\mathcal{I}}_n & \cdots & \sigma_{2p}\boldsymbol{\mathcal{I}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}\boldsymbol{\mathcal{I}}_n & \sigma_{p2}\boldsymbol{\mathcal{I}}_n & \cdots & \sigma_{pp}\boldsymbol{\mathcal{I}}_n \end{bmatrix} \right) \\ &= N_{np}(\boldsymbol{\mu} \otimes \mathbf{1}_n, \boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{I}}_n). \end{aligned}$$

This is immediate from $C(\mathbf{x}_{(i)}, \mathbf{x}_{(j)}) = \sigma_{ij}\boldsymbol{\mathcal{I}}_n$ and $E(\mathbf{x}_{(j)}) = \mu_j \mathbf{1}_n$ and the fact that $\mathbf{X}^v$ is a permutation matrix times the vector on the previous slide (so it's also normal).

Corollary: $\mathbf{X}(n \times p)$ is n.d.m. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow$ $\mathbf{X}^v \sim N_{np}(\boldsymbol{\mu} \otimes \mathbf{1}_n, \boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{I}}_n)$.

prop: $(\mathbf{B}' \otimes \mathbf{A})\mathbf{X}^v = (\mathbf{AXB})^v$.

Proof: First note that

$$(\mathbf{B}' \otimes \mathbf{A})\mathbf{X}^v = \begin{bmatrix} b_{11}\mathbf{A} & b_{21}\mathbf{A} & \cdots & b_{p1}\mathbf{A} \\ b_{12}\mathbf{A} & b_{22}\mathbf{A} & \cdots & b_{p2}\mathbf{A} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1q}\mathbf{A} & b_{2q}\mathbf{A} & \cdots & b_{pq}\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{(1)} \\ \mathbf{x}_{(2)} \\ \vdots \\ \mathbf{x}_{(p)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^p b_{i1}\mathbf{A}\mathbf{x}_{(i)} \\ \sum_{i=1}^p b_{i2}\mathbf{A}\mathbf{x}_{(i)} \\ \vdots \\ \sum_{i=1}^p b_{iq}\mathbf{A}\mathbf{x}_{(i)} \end{bmatrix}.$$

Now let's find the jth column of $\mathbf{A}_{m\times n}\mathbf{X}_{n\times p}\mathbf{B}_{p\times q}$. For *any* $\mathbf{A}_{a\times b}\mathbf{B}_{b\times c}$ the jth column of $\mathbf{AB}$ is $\mathbf{A}\mathbf{b}_{(j)}$. First $\mathbf{AXB} = [\mathbf{A}\mathbf{x}_{(1)} \cdots \mathbf{A}\mathbf{x}_{(p)}]\mathbf{B}$. Thus the jth column of $\mathbf{AXB}$ is $[\mathbf{A}\mathbf{x}_{(1)} \cdots \mathbf{A}\mathbf{x}_{(p)}]\mathbf{b}_{(j)} = \sum_{i=1}^p b_{ij}\mathbf{A}\mathbf{x}_{(i)}$. $\square$

Proof of theorem

$$(\mathbf{B}' \otimes \mathbf{A})\mathbf{X}^v \sim N_{mq}(\underbrace{[\mathbf{B}' \otimes \mathbf{A}][\boldsymbol{\mu} \otimes \mathbf{1}_n]}_{\mathbf{B}'\boldsymbol{\mu} \otimes \mathbf{A}\mathbf{1}_n}, \underbrace{[\mathbf{B}' \otimes \mathbf{A}][\boldsymbol{\Sigma} \otimes \mathcal{I}_n][\mathbf{B}' \otimes \mathbf{A}]'}_{\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B} \otimes \mathbf{A}\mathbf{A}'}).$$

This uses $[\mathbf{A} \otimes \mathbf{B}][\mathbf{C} \otimes \mathbf{D}] = \mathbf{AC} \otimes \mathbf{BD}$ and $[\mathbf{A} \otimes \mathbf{B}]' = \mathbf{A}' \otimes \mathbf{B}'$.

Go back to the theorem, this implies it.

In particular, if $\mathbf{Y} = \mathbf{XB}$ then $\mathbf{Y}$ is d.m. from $N_q(\mathbf{B}'\boldsymbol{\mu}, \mathbf{B}'\boldsymbol{\Sigma}\mathbf{B})$, as $\mathbf{A} = \mathcal{I}_n$.

thm: $\mathbf{X}$ is d.m. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{Y} = \mathbf{AXB}$, $\mathbf{Z} = \mathbf{CXD}$, then $\mathbf{Y}$ indep. of $\mathbf{Z}$ $\Leftrightarrow$ either (a) $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{D} = \mathbf{0}$ or (b) $\mathbf{AC}' = \mathbf{0}$.

You will prove this in your homework, see 3.3.5 (p.88).

Corollary: Let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$ of dimensions $n \times k$ and $n \times (p-k)$. Then $\mathbf{X}_1$ indep. $\mathbf{X}_{2.1} = \mathbf{X}_2 - \mathbf{X}_1 \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$, $\mathbf{X}_1$ d.m. from $N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{X}_{2.1}$ d.m. from $N_{p-k}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1})$ where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$.

Proof: $\mathbf{X}_1 = \mathbf{XB}$ where $\mathbf{B}' = [\mathcal{I}_k \mathbf{0}]$ and $\mathbf{X}_{2.1} = \mathbf{XD}$ where $\mathbf{D}' = [-\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathcal{I}_{p-k}]$. Now use above theorem. $\square$.

Corollary: $\bar{\mathbf{x}}$ indep. $\mathbf{S}$.

Proof: Taking $\mathbf{A} = \frac{1}{n}\mathbf{1}_n'$ and $\mathbf{C} = \mathbf{H} = \mathcal{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n'$ in the theorem gives $\bar{\mathbf{x}}$ indep. $\mathbf{HX}$. $\square$.

Wishart distribution

Note that $\mathbf{S} = \mathbf{X}'[\frac{1}{n}\mathbf{H}]\mathbf{X}$. Quadratic functions of the form $\mathbf{X}'\mathbf{C}\mathbf{X}$ are an ingredient in many multivariate test statistics.

<u>def'n</u>: $M(p \times p) = \mathbf{X}'\mathbf{X}$ where $\mathbf{X}(m \times p)$ is a d.m. from $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ has a Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and d.f. m.
Shorthand: $\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$.

Note that the ijth element of $\mathbf{X}'\mathbf{X}$ is simply $\mathbf{x}'_{(i)}\mathbf{x}_{(j)} = \sum_{k=1}^m x_{ki}x_{kj}$.
The ijth element of $\mathbf{x}_k\mathbf{x}'_k$ is $x_{ki}x_{kj}$. Therefore $\mathbf{X}'\mathbf{X} = \sum_{k=1}^m \mathbf{x}_k\mathbf{x}'_k$.

$$\text{Then } E(\mathbf{M}) = \underbrace{E\left[\sum_{k=1}^{m} \mathbf{x}_k\mathbf{x}'_k\right]}_{E(\mathbf{x}_k)=\mathbf{0}} = \sum_{k=1}^m \boldsymbol{\Sigma} = m\boldsymbol{\Sigma}.$$

Quadratic form involving Wishart

thm: Let $\mathbf{B} \in \mathbb{R}^{p \times q}$ and $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$. Then $\mathbf{B}'\mathbf{M}\mathbf{B} \sim W_q(\mathbf{B}'\mathbf{\Sigma}\mathbf{B}, m)$.

Proof: Let $\mathbf{Y} = \mathbf{X}\mathbf{B}$. Result 3 slides back gives us $\mathbf{Y}$ is d.m. from $N_q(\mathbf{0}, \mathbf{B}'\mathbf{\Sigma}\mathbf{B})$. Then def'n Wishart tells us $\mathbf{Y}'\mathbf{Y} = \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{B}'\mathbf{M}\mathbf{B} \sim W_q(\mathbf{B}'\mathbf{\Sigma}\mathbf{B}, m)$. $\square$

Simple results that follow this theorem

Corollary: Diagonal submatrices of $\mathbf{M}$ (square matrices that share part of a diagonal with $\mathbf{M}$) have a Wishart distribution.

Corollary: $m_{ii} \sim \chi_m^2 \sigma_{ii}$.

Corollary: $\boldsymbol{\Sigma}^{-1/2}\mathbf{M}\boldsymbol{\Sigma}^{-1/2} \sim W_p(\mathcal{I}_p, m)$.

Corollary: $\mathbf{M} \sim W_p(\mathcal{I}_p, m)$ and $B(p \times q)$ s.t. $\mathbf{B}'\mathbf{B} = \mathcal{I}_q$ then $\mathbf{B}'\mathbf{M}\mathbf{B} \sim W_q(\mathcal{I}_q, m)$.

Corollary: $\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$ and $\mathbf{a}$ s.t. $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \neq \mathbf{0} \Rightarrow \frac{\mathbf{a}'\mathbf{M}\mathbf{a}}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \sim \chi_m^2$.

All use different $\mathbf{B}$ in the theorem on the previous slide plus minor manipulation.

Wisharts add!

thm $\mathbf{M}_1 \sim W_p(\mathbf{\Sigma}, m_1)$ indep. $\mathbf{M}_2 \sim W_p(\mathbf{\Sigma}, m_2) \Rightarrow$
$\mathbf{M}_1 + \mathbf{M}_2 \sim W_p(\mathbf{\Sigma}, m_1 + m_2)$.

Proof: Let $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$. Then $\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{X}'\mathbf{X}$. Now use the def'n of Wishart. $\square$

We are just adding m_2 more independent rows onto $\mathbf{X}_1$.

Cochran's theorem

thm: If $\mathbf{X}(n \times p)$ d.m. from $N_p(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{C}(n \times n)$ is symmetric w/ eigenvalues $\lambda_1, \dots, \lambda_n$ then

(a) $\mathbf{X}'\mathbf{C}\mathbf{X} \overset{D}{=} \sum_{i=1}^n \lambda_i \mathbf{M}_i$ where $\mathbf{M}_1, \dots, \mathbf{M}_n \overset{iid}{\sim} W_p(\mathbf{\Sigma}, 1)$.

(b) $\mathbf{X}'\mathbf{C}\mathbf{X} \sim W_p(\mathbf{\Sigma}, r) \Leftrightarrow \mathbf{C}$ idempotent where $r = \text{tr}\mathbf{C} = \text{rank}\mathbf{C}$.

(c) $n\mathbf{S} \sim W_p(\mathbf{\Sigma}, n-1)$.

Proof The spectral decomposition of $\mathbf{C}$ is $\mathbf{C} = [\gamma_1 \cdots \gamma_n]\mathbf{\Lambda}[\gamma_1 \cdots \gamma_n]' = \sum_{i=1}^n \lambda_i \gamma_i \gamma_i'$. Then $\mathbf{X}'\mathbf{C}\mathbf{X} = \sum_{i=1}^n \lambda_i [\mathbf{X}'\gamma_i][\mathbf{X}'\gamma_i]' = \sum_{i=1}^n \lambda_i [\gamma_i'\mathbf{X}]'[\gamma_i'\mathbf{X}]$. General transformation theorem ($\mathbf{A} = \gamma_i'$ & $\mathbf{B} = \mathcal{I}_p$) tells us that $\gamma_i'\mathbf{X}$ is d.m. from $N_p(\mathbf{0}, \mathbf{\Sigma})$ so (a) follows from def'n Wishart. Part (b): $\mathbf{C}$ idempotent $\Rightarrow$ there are r $\lambda_i = 1$ and $n - r$ $\lambda_i = 0$, hence tr $\mathbf{C} = \lambda_1 + \cdots \lambda_n = r$. Now use part (a). For part (c) note that $\mathbf{H}$ is idempotent and rank $n-1$. $\square$

This is a biggie. Lots of stuff that will be used later.

Drum roll...

If $\mathbf{x}_1, \ldots, \mathbf{x}_n \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \tfrac{1}{n}\boldsymbol{\Sigma}),$$

$$n\mathbf{S} \sim W_p(\boldsymbol{\Sigma}, n-1),$$

and $\bar{\mathbf{x}}$ indep. of $\mathbf{S}$.

This is a generalization of the univariate $p = 1$ case where $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ indep. of $ns^2 \sim \sigma^2\chi^2_{n-1}$. This latter result is used to cook up a t_{n-1} distribution:

$$\frac{\bar{x} - \boldsymbol{\mu}}{\sqrt{s^2/n}} \sim t_{n-1},$$

by def'n. We'll shortly generalize this to p dimensions, but first one last result.

Here is Craig's theorem.

<u>thm</u> $\mathbf{X}$ d.m. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{C}_1, \ldots, \mathbf{C}_k$ are symmetric, then $\mathbf{X}'\mathbf{C}_1\mathbf{X}, \ldots, \mathbf{X}'\mathbf{C}_k\mathbf{X}$ are indep. if $\mathbf{C}_r\mathbf{C}_s = \mathbf{0}$ for all $r \neq s$.

Proof: Let's do it for two projection matrices. Write $\mathbf{X}'\mathbf{C}_1\mathbf{X} = \mathbf{X}'\mathbf{M}_1\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1\mathbf{M}_1'\mathbf{X}$ and $\mathbf{X}'\mathbf{C}_2\mathbf{X} = \mathbf{X}'\mathbf{M}_2\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2\mathbf{M}_2'\mathbf{X}$. Note that $\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i$ as the e-values are either 1 or 0. Theorem (slide 14) says $\boldsymbol{\Lambda}_1\mathbf{M}_1'\mathbf{X}$ indep. $\boldsymbol{\Lambda}_2\mathbf{M}_2'\mathbf{X} \Leftrightarrow$
$[\boldsymbol{\Lambda}_1\mathbf{M}_1'][\boldsymbol{\Lambda}_2\mathbf{M}_2']' = \boldsymbol{\Lambda}_1\mathbf{M}_1'\mathbf{M}_2\boldsymbol{\Lambda}_2 = \mathbf{0}$. But
$\mathbf{0} = \mathbf{C}_1\mathbf{C}_2 = \mathbf{M}_1\boldsymbol{\Lambda}_1\mathbf{M}_1'\mathbf{M}_2\boldsymbol{\Lambda}_2\mathbf{M}_2' \Rightarrow \boldsymbol{\Lambda}_1\mathbf{M}_1'\mathbf{M}_2\boldsymbol{\Lambda}_2 = \mathbf{0}$. $\square$

This will come in handy in finding the sampling distribution of common test statistics under H_0.

Hotelling's T^2

Recall, using obvious notation, $\frac{N(0,1)}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu$. Used for one and two-sample t tests for univariate outcomes. We'll now generalize this distribution.

<u>def'n</u>: Let $\mathbf{d} \sim N_p(\mathbf{0}, \mathcal{I}_p)$ indep. $\mathbf{M} \sim W_p(\mathcal{I}_p, m)$. Then $m\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} \sim T^2(p, m)$.

<u>thm</u>: Let $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indep. $\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$. Then $m(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim T^2(p, m)$.

Proof: Take $\mathbf{d}^* = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ and $\mathbf{M}^* = \boldsymbol{\Sigma}^{-1/2}\mathbf{M}\boldsymbol{\Sigma}^{-1/2}$ and use def'n of T^2. □

<u>Corollary</u>: $\bar{\mathbf{x}}$ and $\mathbf{S}$ are sample mean and covariance from $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow$
$(n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim T^2(p, n-1)$.

Proof: Substitute $\mathbf{M} = n\mathbf{S}$, $m = n-1$, and $\mathbf{x} - \boldsymbol{\mu}$ for $\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu})$ in the theorem above. □

Hotelling's T^2 is a scaled F

<u>thm</u>: $T^2(p,m) = \frac{mp}{m-p+1} F_{p,m-p+1}$.

To prove this we need some ingredients...

Let $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$ and take $\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}$ where $\mathbf{M}_{11} \in \mathbb{R}^{a \times a}$ and $\mathbf{M}_{22} \in \mathbb{R}^{b \times b}$ and $a + b = p$. Further, let $\mathbf{M}_{22.1} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$.

thm: Let $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$ where $m > p$. Then $\mathbf{M}_{22.1} \sim W_b(\mathbf{\Sigma}_{22.1}, m - a)$.

Proof: Let $\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2]$, so

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}.$$

Then

$$\mathbf{M}_{22.1} = \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X}_2 = \mathbf{X}_2'\mathbf{P}\mathbf{X}_2 = \mathbf{X}_{2.1}'\mathbf{P}\mathbf{X}_{2.1},$$

where $\mathbf{P} = \mathcal{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1$ is o.p. matrix onto $\mathcal{C}(\mathbf{X}_1)^{\perp}$ and $\mathbf{X}_{2.1}|\mathbf{X}_1 = \mathbf{X}_2 - \mathbf{X}_1\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$. Theorem on slide 14 tells us $\mathbf{X}_{2.1}$ is d.m. from $N_b(\mathbf{0}, \mathbf{\Sigma}_{22.1})$ (not dim. p as in the book). So Cochran's theorem tells us $\mathbf{M}_{22.1}|\mathbf{X}_1 \sim W_b(\mathbf{\Sigma}_{22.1}, m - a)$. This doesn't depend on $\mathbf{X}_1$ so it's the marginal dist'n as well. $\square$

Proof, Hotelling's T^2 is a scaled F

lemma: If $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$, $m > p$ then $\frac{1}{[\mathbf{M}^{-1}]_{pp}} \sim \frac{1}{[\mathbf{\Sigma}^{-1}]_{pp}} \chi^2_{m-p-1}$.

Proof: In general, for partitioned matrices,

$$\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})^{-1} & -\mathbf{M}_{11}^{-1}\mathbf{M}_{12}(\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12})^{-1} \\ -\mathbf{M}_{22}^{-1}\mathbf{M}_{21}(\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})^{-1} & (\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12})^{-1} \end{bmatrix}.$$

Now let $\mathbf{M}_{11}$ be upper left $(p-1) \times (p-1)$ submatrix of $\mathbf{M}$ and m_{22} the lower right 1×1 "scalar matrix." Then, where $\sigma_{22.1} = \frac{1}{[\mathbf{\Sigma}^{-1}]_{pp}}$,

$\frac{1}{[\mathbf{M}^{-1}]_{pp}} = \frac{1}{1/m_{22.1}} = m_{22.1} \sim W_1(\sigma_{22.1}, m-(p-1)) = \sigma_{22.1}\chi^2_{m-p-1}$. $\square$

---

thm: If $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$, $m > p$ then $\frac{\mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{M}^{-1}\mathbf{a}} \sim \chi^2_{m-p+1}$.

Proof: Let $\mathbf{A} = [\mathbf{a}_{(1)} \cdots \mathbf{a}_{(p-1)}\mathbf{a}]$. Then
$\mathbf{N} = \mathbf{A}^{-1}\mathbf{M}(\mathbf{A}^{-1})' \sim W_p(\mathbf{A}^{-1}\mathbf{\Sigma}(\mathbf{A}^{-1})', m)$. So

$$\frac{1}{[\mathbf{N}^{-1}]_{pp}} = \frac{1}{[\mathbf{A}\mathbf{M}^{-1}\mathbf{A}']_{pp}} = \frac{1}{\mathbf{a}'\mathbf{M}^{-1}\mathbf{a}} \sim \frac{1}{\mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{a}}\chi^2_{m-p+1}.$$

Noting that the ppth element of $[\mathbf{A}^{-1}\mathbf{\Sigma}(\mathbf{A}^{-1})']^{-1}$ is $\frac{1}{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}$. $\square$

Proof, Hotelling's T^2 is a scaled F

Recall $m\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} \sim T^2(p,m)$ where $\mathbf{d} \sim N_p(\mathbf{0}, \mathcal{I}_p)$ indep. of $\mathbf{M} \sim W_p(\mathcal{I}_p, m)$. Given $\mathbf{d}$, $\beta = \frac{\mathbf{d}'\mathbf{d}}{\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}} \sim \chi^2_{m-p+1}$ (last slide). Since this is independent of $\mathbf{d}$, β indep. $\mathbf{d}$ and this is the marginal dist'n as well.

$$m\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} = \frac{m\mathbf{d}'\mathbf{d}}{\mathbf{d}'\mathbf{d}/\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}} = m\frac{\chi^2_p}{\chi^2_{m-p+1}} = \frac{mp}{m-p+1}F_{p,m-p+1}.\square$$

Corollary: $\bar{\mathbf{x}}$ and $\mathbf{S}$ are sample mean and covariance from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\frac{n-p}{p}(\bar{\mathbf{x}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \sim F_{p,n-p}$.

Two more distributional results

Corollary: $|\mathbf{M}|/|\mathbf{M}+\mathbf{dd}'| \sim B(\frac{1}{2}(m-p+1), \frac{p}{2})$.

Proof: For $\mathbf{B}(p \times n)$ and $\mathbf{C}(n \times p)$, $|\mathcal{I}_p + \mathbf{BC}| = |\mathcal{I}_n + \mathbf{CB}|$.
Since $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$, we can write this as
$\frac{1}{|\mathcal{I}_p+\mathbf{M}^{-1}\mathbf{dd}'|} = \frac{1}{|\mathcal{I}_1+\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}|} = \frac{1}{1+\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}} = \frac{1}{1+\frac{p}{m-p+1}F_{p,m-p+1}}$. Recall
if $x \sim F_{\nu_1,\nu_2}$ then $\frac{\nu_1 x/\nu_2}{1+\nu_1 x/\nu_2} \sim B(\frac{\nu_1}{2}, \frac{\nu_2}{2})$ and $\frac{1}{1+\nu_1 x/\nu_2} \sim B(\frac{\nu_2}{2}, \frac{\nu_1}{2})$.
□

Corollary: $\mathbf{d} \sim N_p(\mathbf{0}, \mathcal{I}_p)$ indep. $\mathbf{M} \sim W_p(\mathcal{I}_p, m)$ then
$\mathbf{d}'\mathbf{d}(1 + 1/\{\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}\}) \sim \chi^2_{m+1}$ indep. of $\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}$.

Proof: β indep. $\mathbf{d}'\mathbf{d}$ (last slide); both χ^2 so their sum is indep. of their ratio. Sum of two indep. χ^2 is also χ^2; the d.f. add. □

Two-sample Hotelling T^2

Let $D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_u^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ estimate $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ where $\mathbf{S}_u = \frac{1}{n-2}[n_1\mathbf{S}_1 + n_2\mathbf{S}_2]$. Then

<u>thm</u>: Let $\mathbf{X}_1$ d.m. from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ indep. $\mathbf{X}_2$ d.m. from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. If $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ then $\frac{n_1 n_2}{n_1+n_2}D^2 \sim T^2(p, n-2)$.

Proof: $\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 \sim N_p(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \frac{1}{n_1}\boldsymbol{\Sigma}_1 + \frac{1}{n_2}\boldsymbol{\Sigma}_2)$. When $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, $\mathbf{d} \sim N_p(\mathbf{0}, c\boldsymbol{\Sigma})$ where $c = \frac{n_1+n_2}{n_1 n_2}$. Also $\mathbf{M} = n_1\mathbf{S}_1 + n_2\mathbf{S}_2 \sim W_p(\boldsymbol{\Sigma}, n_1 + n_2 - 2)$ as independent Wisharts w/ same scale add; $c\mathbf{M} \sim W_p(c\boldsymbol{\Sigma}, n_1 + n_2 - 2)$. $\mathbf{M}$ indep. $\mathbf{d}$ as $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S}_1, \mathbf{S}_2$ mutually indep. Stirring all ingredients together gives $\frac{D^2}{c} = (n-2)\mathbf{d}'(c\mathbf{M})^{-1}\mathbf{d} \sim T^2(p, n-2)$. $\square$.

Generalization of F statistic

We've already generalized the t for multivariate data; now it's time for the F.

Let $\mathbf{A} \sim W_p(\mathbf{\Sigma}, m)$ indep. of $\mathbf{B} \sim W_p(\mathbf{\Sigma}, n)$ where $m \geq p$. $\mathbf{A}^{-1}$ exists a.s. and we will examine aspects of $\mathbf{A}^{-1}\mathbf{B}$.

Note that this reduces to the ratio of indep, χ^2 in the univariate $p = 1$ case.

lemma: $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$, $m \geq p$, $\Rightarrow |\mathbf{M}| = |\mathbf{\Sigma}| \prod_{i=0}^{p-1} \chi^2_{m-i}$.

Proof: By induction. For $p = 1$ $|\mathbf{M}| = m \sim \sigma^2 \chi^2_m$. For $p > 1$ let $\mathbf{M}_{11}$ be upper left $(p-1) \times (p-1)$ submatrix of $\mathbf{M}$ and m_{22} the lower right 1×1 "scalar matrix" (slide 24). The induction hypothesis says $|\mathbf{M}_{11}| = |\mathbf{\Sigma}_{11}| \prod_{i=0}^{p-2} \chi^2_{m-i}$. Slide 24 implies that $m_{22.1}$ indep. $\mathbf{M}_{11}$ and $m_{22.1} \sim \sigma_{22.1} \chi^2_{m-p+1}$. The result follows by noting that $|\mathbf{M}| = |\mathbf{M}_{11}| m_{22.1}$ and $|\mathbf{\Sigma}| = |\mathbf{\Sigma}_{11}| \sigma_{22.1}$ (p. 457 or expansion of determinant using cofactors). $\square$

thm: Let $\mathbf{A} \sim W_p(\mathbf{\Sigma}, m)$ indep. of $\mathbf{B} \sim W_p(\mathbf{\Sigma}, n)$ where $m \geq p$ and $n \geq p$. Then $\phi = |\mathbf{B}|/|\mathbf{A}| \propto \prod_{i=1}^{p} F_{n-i+1, m-i+1}$.

Proof: Using the lemma

$$\phi = \prod_{i=0}^{p} \frac{\chi^2_{n-i}}{\chi^2_{m-i}} = \prod_{i=0}^{p} \frac{n-i}{m-i} F_{n-i+1, m-i+1}. \square$$

Wilk's lambda

Wilk's lambda, a generalization of the beta variable, appears later on when performing LRT:

def'n: $\mathbf{A} \sim W_p(\mathcal{I}_p, m)$ indep. $\mathbf{B} \sim W_p(\mathcal{I}_p, n)$ and $m \geq p$

$$\Lambda = |\mathbf{A}|/|\mathbf{A} + \mathbf{B}| \sim \Lambda(p, m, n),$$

has a Wilk's lambda distribution with parameters (p, m, n)

thm: $\Lambda \sim \prod_{i=1}^{n} u_i$ where $u_1, \ldots, u_n$ are mutually independent and $u_i \sim B(\frac{1}{2}(m + i - p), \frac{p}{2})$.

Proof Wilk's lambda in product of betas

Let $\mathbf{X}(n \times p)$ be d.m. from $N_p(\mathbf{0}, \mathcal{I}_p)$, $\mathbf{B} = \mathbf{X}'\mathbf{X}$ and $\mathbf{X}_i$ be first i rows of $\mathbf{X}$. Let $\mathbf{M}_i = \mathbf{A} + \mathbf{X}_i'\mathbf{X}_i$. Then $\mathbf{M}_0 = \mathbf{A}$, $\mathbf{M}_n = \mathbf{A} + \mathbf{B}$, and $\mathbf{M}_i = \mathbf{M}_{i-1} + \mathbf{x}_i\mathbf{x}_i'$. Then

$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A} + \mathbf{B}|} = \prod_{i=1}^{n} \frac{|\mathbf{M}_{i-1}|}{|\mathbf{M}_i|} = \prod_{i=1}^{n} u_i.$$

Corollary on slide 27 implies $u_i \sim B(\frac{1}{2}(m + i - p), \frac{p}{2})$.

The independence part takes some work...

Characterization of independence of matrix & vector

lemma: Let $\mathbf{W} \in \mathbb{R}^{p \times p}$ and $\mathbf{x} \in \mathbb{R}^p$. If $\mathbf{x}$ is indep. of $(\mathbf{g}_1'\mathbf{W}\mathbf{g}_1, \ldots, \mathbf{g}_p'\mathbf{W}\mathbf{g}_p)$ for all orthogonal $\mathbf{G} = [\mathbf{g}_1 \cdots \mathbf{g}_p]'$ then $\mathbf{x}$ indep. $\mathbf{W}$.

Proof: The c.f. of $\{2^{I\{i<j\}} w_{ij} : i \le j\}$ is $E\{e^{i\text{tr}(\mathbf{WT})}\}$ where $\mathbf{T}$ is symmetric. The c.f. of $(\mathbf{x}, \mathbf{W})$ is thus characterized by $\phi_{\mathbf{W},\mathbf{x}}(\mathbf{T}, \mathbf{s}) = E\{e^{i\text{tr}(\mathbf{WT})} e^{i\mathbf{s}'\mathbf{x}}\}$. If $\mathbf{x}$ indep. $\text{tr}\mathbf{WT}$ for all symmetric $\mathbf{T}$ then the c.f. factors and $\mathbf{x}$ indep. $\mathbf{W}$.
Let $\mathbf{A} = \mathbf{G\Lambda G}' = \sum_{i=1}^p \lambda_i \mathbf{g}_i \mathbf{g}_i'$ be spectral decomposition. Then

$$\text{tr}\mathbf{AW} = \text{tr}\left\{\sum_{i=1}^p \lambda_i \mathbf{g}_i \mathbf{g}_i' \mathbf{W}\right\} = \sum_{i=1}^p \lambda_i \underbrace{\mathbf{g}_i'\mathbf{W}\mathbf{g}_i}_{\mathbf{x}\text{ indep. these}} \quad . \square$$

Showing independence of $u_1, \ldots, u_n$, continued...

thm: $\mathbf{d} \sim N_p(\mathbf{0}, \mathcal{I}_p)$ indep. $\mathbf{M} \sim W_p(\mathcal{I}_p, m)$ then
$\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} \sim \frac{p}{m-p+1} F_{p,m-p+1}$ indep. $\mathbf{M} + \mathbf{dd}' \sim W_p(\mathcal{I}_p, m+1)$.

Proof: Let $\mathbf{G} = [\mathbf{g}_1 \cdots \mathbf{g}_p]'$ orthogonal matrix and take
$\mathbf{X}((m+1) \times p) = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{x}'_{m+1} \end{bmatrix}$. Here $\mathbf{M} = \mathbf{X}'_1\mathbf{X}_1$ and $\mathbf{d} = \mathbf{x}_{m+1}$. Let

$$\mathbf{Y} = \mathbf{XG}' = [\mathbf{Xg}_1 \cdots \mathbf{Xg}_p] = [\mathbf{Y}_{(1)} \ldots \mathbf{Y}_{(p)}].$$

Then $q_j = \mathbf{g}'_j[\mathbf{M} + \mathbf{dd}']\mathbf{g}_j = \mathbf{g}'_j\mathbf{X}'\mathbf{X}\mathbf{g}_j = ||\mathbf{Y}_{(j)}||^2$. Since
$\mathbf{Y}^v \sim N_{np}(\mathbf{0}, \mathcal{I}_{np})$, $\mathbf{Y}^v$ is spherically symmetric. Define
$h(\mathbf{Y}) = \mathbf{y}'_{m+1}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{y}_{m+1} = \mathbf{d}'\mathbf{M}^{-1}\mathbf{d}$ and note that
$h(\mathbf{Y}) = h(\mathbf{YD})$ for all diagonal $\mathbf{D}$. Theorem on p. 48 implies q_j
indep. $h(\mathbf{Y})$ for $j = 1, \ldots, p$. Now use lemma on previous slide. $\square$

Showing independence of $u_1, \ldots, u_n$, continued...

Theorem on last slide implies $\frac{1}{u_i} = |\mathbf{M}_i|/|\mathbf{M}_{i-1}| = 1 + \mathbf{x}_i'\mathbf{M}_{i-1}^{-1}\mathbf{x}_i$ indep. $\mathbf{M}_i$. Finally,

$$\mathbf{M}_{i+j} = \mathbf{M}_i + \sum_{k=1}^{j} \underbrace{\mathbf{x}_{i+k}\mathbf{x}_{i+k}'}_{u_i \text{ indep. of}} ,$$

so for any i, u_i indep. of $u_{i+1}, \ldots, u_n$. $\square$

Two last results...

When m is large, can also use Bartlett's approximation:

$$-\{m - \tfrac{1}{2}(p - n + 1)\} \log \Lambda(p, m, n) \overset{\bullet}{\sim} \chi^2_{np}.$$

---

def'n: $\mathbf{A} \sim W_p(\mathcal{I}_p, m)$ indep. $\mathbf{B} \sim W_p(\mathcal{I}_p, n)$ and $m \geq p$.
$\theta(p, m, n)$, the largest eigenvalue of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$ is called the greatest root statistic with parameters (p, m, n).

MKB (p. 84) gives relationships between $\Lambda(p, m, n)$ and $\theta(p, m, n)$