B. The Multivariate Normal Distribution

B1. Definitions and First Results

Standard and general p-dim normal characteristic function, basic properties, spherical symmetry, partitions, best linear predictor, repeated sampling, basic statistics, results, Kronecker product, and matrix norms, properties, rotational invariance, removing means, prolate data example: Spherical symmetry and the t-test, Benjamin-Hochberg procedure.

B2. Normal matrices, Triangular Coordinates, and Hsu's Lemma: Jacobian to triangular coordinates, $J(x \to T)$, normal-$\chi^2$ density formula for $T$, $T(x \to T, \Lambda)$, Hsu's Lemma.

B3. The Wishart Distribution

Wishart density, properties and facts, removing the means, decomposition property, $\mathbb{E}[\mathbf{S}] = \mathbf{V}$, Functions of a Wishart matrix: $S$, $R$, delta method for Jacobians, Moments of the Wishart correlation Wishart moments, $\text{Re} \phi$.

Ist Carolio Misra's example.

B4. Hotelling's $T^2$ and the Mahalanobis Distance

1-sample $t$-test, one-sample $T^2$ test, geometric interpretation, null distribution, data-mapping example, Two-sample $T^2$ test, geometric interpretation, Normal null distribution Hotelling's Theorem, Mahalanobis distance, partitioning $\Lambda$. 
### B. The Multivariate Normal Distribution

#### B.1. Definitions and First Results

**Standard Normal** $Z \sim \mathcal{N}(0, I_p) = \begin{pmatrix} \frac{z_1}{2} \\ \vdots \\ \frac{z_p}{2} \end{pmatrix}$ if $Z \sim \mathcal{N}(0, I)$:

$$f(Z) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2} Z^T Z} = e^{-\frac{1}{2} \|z\|^2} / (2\pi)^{\frac{p}{2}}$$

### Characteristic Function

By completing the square in the exponent we get

$$E \left\{ e^{iZ^T Z} \right\} = e^{\frac{1}{2} \|a\|^2}$$

so choosing $a = it$ give characteristic function

$$\Phi_Z(t) = E \left\{ e^{iZ^T Z} \right\} = e^{\frac{1}{2} \|t\|^2}$$

### General Normal

Let $\mu$ be $p \times 1$ vector, $\Sigma^{\frac{1}{2}}$ a $p \times p$ matrix. Then

$$X = \mu + \Sigma^{\frac{1}{2}} Z \sim \mathcal{N}(\mu, \Sigma)$$

$$\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}'$$

Properties and Facts

(i) \( E(X) = \mu \), \( \text{Var}(X) = \Sigma \) \[\text{From } E(Z = 0), \text{Var}(Z) = I, \]\n\[X \sim (\mu, \Sigma)\]

(ii) If \( \text{rank}(\Sigma) = p \) then \( \Sigma > 0 \) and
\[
\Phi(x) = \frac{e^{-\frac{1}{2}x^T \Sigma^{-1} x}}{\sqrt{2\pi}^p \sqrt{\det(\Sigma)}} \quad \left[ J(x \rightarrow x) = \frac{\partial \Phi}{\partial x} \right]^{-\frac{1}{2}}
\]

(iii) \( \psi_X(t) = e^{it\mu - \frac{1}{2}t^2 \Sigma} \quad \left[ = \text{E} e^{itX} \right] \]

(iv) For any choice of \( \mu \) and \( \Sigma > 0 \), there is a unique \( \mathcal{N}_p(\mu, \Sigma) \) distribution \[\text{since } \psi_X(t) = \psi_X(4i) \]

(v) \( X \sim \mathcal{N}_p(\mu, \Sigma) \) \[\text{wp } 1 \quad \Gamma(X - \mu \in \mathcal{K}_c) = \mathcal{L}_c(\mathcal{K}_c, \Sigma) \text{, subsec A.3} \]

(vi) \( Y = \mathcal{N}(\mu + \Delta X, \Sigma) \quad \text{wp } \frac{1}{2} \quad \Gamma(Y \in \mathcal{K}_c \mid \Delta X) = \mathcal{L}_c(\mathcal{K}_c, \Sigma) \text{, subsec A.3} \]

(vii) \( "\text{Kurtosis}\) \quad \text{E}(X - \mu)_2^2 - \frac{1}{2} X^T \Sigma^{-1} X \quad \left[ = \sum \sigma_i^2 \right] \quad \text{Skewness} = 0 \quad \text{[differentiation, } \psi_X(t) \text{.]} \]
The standard normal is spherically symmetric.

\[ X = \mathbb{P}^{1/2} \sim \mathbb{P}(0, I) \sim \mathbb{Z} \]

and \( Y = \mathbb{P}X \sim \mathbb{P}(0, I) \): "The spherically normal projects into a lower dimensional spherically normal".

The unit vector \( U = \frac{X}{\|X\|} \) is uniformly distributed on the surface.
of the unit sphere in $\mathbb{R}^p$, independently of $|x|^2 \sim \chi_p^2$.

**Hausdorff 2**

(a) Use multivariate polar coordinates to show the independence. (b) How is the uniform density on the surface of the unit sphere in $\mathbb{R}^p$ expressed in terms of $(\theta_1, \theta_2, \ldots, \theta_p)$?

(c.i) If $(X) \sim \mathcal{N}_{p,q} \left( \mu, \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \right)$ then

   $$X \sim \mathcal{N}_p \left( \mu, \mathbf{V}_{xx} \right)$$

$$[\mathbf{L}_x(t) = \mathbf{Y}_{x,t} (t,0)]$$

(c.ii) If $\mathbf{r} = 0$ then $X \perp Y$

$$[\mathbf{C}_{xy}(t,\tau) = \mathbf{C}_x(t) \mathbf{C}_y(\tau)]$$

(c.iii) \[ Y \mid X \sim \mathcal{N} \left( \nu + \mathbf{r}_y \mathbf{r}_x^{-1} (X - \mu), \mathbf{V}_{y|x}^{-1} \right) \]

$$[\mathbf{C}_Y(t) = \mathbf{C}_X(t) \mathbf{C}_Y]$$

**Hausdorff 2** Verify using characteristic function.
- From the general theory of Section A.2:
  - \( \hat{Y} = \tilde{Y} + \frac{1}{\sigma^2} G' (X - \mu) \) is best linear predictor of \( Y \) from \( X \)
  - \( \text{Cov}(\hat{Y}) = \frac{1}{\sigma^2} \frac{G'}{G} \frac{G'}{G} \frac{G'}{G} \)
  - \( \hat{Y} = Y - \bar{Y} \) \( \text{Cov}(\tilde{Y}) = \frac{1}{\sigma^2} \frac{G'}{G} \frac{G'}{G} \frac{G'}{G} \)
  - \( \text{Cov}(X, \hat{Y}) = 0 \)

(xiv) For the multivariate normal, the best linear predictor is the best predictor and the residual covariance after linear prediction equates the conditional covariance \( \text{Cov}(Y | X) \).

\[
Y = \hat{Y} + \frac{1}{\sigma^2} G' (X - \mu) \quad \text{when} \quad \hat{Y} \text{ linear in } X \quad \text{and} \quad \hat{Y} \sim N \left( 0, \frac{1}{\sigma^2} \frac{G'}{G} \frac{G'}{G} \right) \mid X
\]
Homework 3 / Multivariate Cauchy \( X \sim \mathcal{N}_p(0, \mathbb{I}) \| z \sim \mathcal{N}(0, I) \).

Show that \( Y = X/z \) has characteristic function \( \phi_Y(t) = e^{-(4t^2 + t^2 + 1)} \).

And density
\[
\phi_Y(y) = \frac{1}{\pi^{p+1}} \left[ 1 + y'y \right]^{-\frac{p+1}{2}}.
\]

Repeated Sampling

Now suppose we observe \( n \) independent replications of \( X \sim \mathcal{N}_p(\mu, \Sigma) \),

\[
X_1, X_2, \ldots, X_n \sim \mathcal{N}_p(\mu, \Sigma) \quad \Sigma > 0.
\]

Let \( \bar{X} = \left( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \right) \), observed \( n/\mu \), \( \bar{x} \).

For
\[
\int_{\mathbb{R}^n} f_{\bar{X}}(x) \, dx = (2\pi)^{-n/2} \mathbb{I}^{-1/2} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_i)^2} \mathbb{I}^{-1/2}.
\]
Using
\[ \sum_{\lambda=0}^{\lambda_0} \mathbb{E} (x-\mu) \mathbb{E} (x-\mu) = \mathbb{E} \left( \sum_{\lambda=0}^{\lambda_0} (x-\mu)^2 \right) \]

we get
\[ \mathbb{E} \left[ \sum_{\lambda=0}^{\lambda_0} (x-\bar{x}) (x-\bar{x}) + n(\bar{x}-\mu)^2 \right] \]

\[ = \mathbb{E} (\bar{x}^2) + n(\bar{x}-\mu)^2 \mathbb{E} (\bar{x}-\mu) \]

\[ f_m, \mathbb{E} (x) = (2\pi)^{-\frac{m}{2}} \mathbb{E} \left[ e^{-\frac{1}{2} \sum_{\lambda=0}^{\lambda_0} (x-\bar{x})^2 + n(\bar{x}-\mu)^2} \right] \]

Hence \( \frac{1}{\lambda_0} \) Show that we can also write

\[ f_m, \mathbb{E} (x) = (2\pi)^{-\frac{m}{2}} \mathbb{E} \left[ e^{-\frac{1}{2} \sum_{\lambda=0}^{\lambda_0} (x-\bar{x})^2 + n(\bar{x}-\mu)^2} \right] \]

where
\[ \sum_{i=0}^{\lambda_0} \text{element of } S \]

\[ \sum_{i=0}^{\lambda_0} \text{element of } S \]

\[ \sum_{i=0}^{\lambda_0} \text{element of } S \]

\[ \sum_{i=0}^{\lambda_0} \text{element of } S \]

\[ \delta_{ij} = 1 \text{ if } i = j \text{ or } \neq 0 \]
- Define $z$ as the vector of elements $x_i$, $i = 1$.

- Basic Statistical Theory gives

(i) $\hat{\theta}(x, z)$ is sufficient for $(\mu, \sigma^2)$

(ii) $f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (parametric exponential family)

<table>
<thead>
<tr>
<th>Natural Parameter</th>
<th>Sufficient Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{x}$, $x_i$, $i \geq 1$</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$\frac{1}{n} \sum (x_i - \bar{x})^2$, $i \geq 1$</td>
</tr>
</tbody>
</table>

- (v) Number of parameters does not depend on $n$.

- Therefore $(\bar{x}, \hat{\sigma})$ minimally sufficient and complete.

- For any parameter $\theta = (\mu, \sigma^2)$, there is an unbiased estimate $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$, and if there is one, it is UMVU.
Definition: \( \hat{\theta} \) unbiased \( \text{in UMVU for } \theta \) if, for any other unbiased estimate \( \tilde{\theta} \), \( \text{Cov}(\hat{\theta}) - \text{Cov}(\tilde{\theta}) \geq 0 \).

Lemma: \( \hat{\theta} \text{ UMVU for } \theta \Rightarrow A \hat{\theta} \text{ UMVU for } A \theta \) (and so \( \tilde{\theta} \text{ UMVU for } \tilde{\theta} \)).

Examples:
(a) \( \bar{X} \text{ UMVU for } \mu \) \( \frac{\sum_{i=1}^{n} x_i}{n} \text{ UMVU for } \mu \)

Exercise 5/24: Find U MVU estimates for (a) For \( \mu \).
(b) \( \tilde{\mu} \).
(c) It can be shown that for \( \mu \), \( \tilde{\mu} = (\bar{X}, \bar{Y}) \), \( E(\hat{\theta}) = \tilde{\theta} \). Show that for any \( \mu \), \( \hat{\theta} = \tilde{\theta} \) is U MVU for \( \tilde{\theta} \).
Kronecker Products and Random Normal Matrices

Independent Sampling $X = (x_1, x_2, \ldots, x_n)$ with $x_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ can be thought of as a single random vector

$$X \sim \mathcal{N}(\mu, \Sigma)$$

with

$$\mu = \left( \begin{array}{c} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{array} \right)$$

and

$$\Sigma = \left( \begin{array}{ccc} \Sigma_{11} & \cdots & \Sigma_{1n} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \Sigma_{nn} \end{array} \right)$$

Linear Transformations We will want to consider transformations

$$Y \sim A X B$$

with $A \in \mathbb{R}^{p \times p}$ and $B \in \mathbb{R}^{q \times q}$.
and it gets very clumsy stringing out coordinates.

**Kronecker Products** Are designed to handle this situation:

\[
A \otimes B = \begin{pmatrix}
B_1 A & B_2 A & \cdots & B_{rA} A \\
B_1 A & B_2 A & \cdots & B_{rA} A \\
\vdots & \vdots & \ddots & \vdots \\
B_1 A & B_2 A & \cdots & B_{rA} A
\end{pmatrix}
\]

**Definition**

\[
P \otimes Q = \begin{pmatrix}
P_1 Q & P_2 Q & \cdots & P_{rP} Q \\
P_1 Q & P_2 Q & \cdots & P_{rP} Q \\
\vdots & \vdots & \ddots & \vdots \\
P_1 Q & P_2 Q & \cdots & P_{rP} Q
\end{pmatrix}
\]

**Warning** The R function 'kronecker' reverses the roles of A and B.

\[
\begin{pmatrix}
(a_{11}, a_{12}) & (a_{21}, a_{22}) \\
\vdots & \vdots \\
(a_{r1}, a_{r2}) & (a_{r1}, a_{r2})
\end{pmatrix}
\]

\[
(a_{11} \cdot a_{12} \cdot \ldots \cdot a_{r1} \cdot a_{r2})
\]

**Facts**

(i) \((A \otimes B)' = A' \otimes B'\)

(ii) \((A \otimes B)' = A' \otimes B'\) if both inverses exist

(iii) \((A \otimes B)(C \otimes D) = AC \otimes BD\)

(iv) \(Y = (A \times B) = (A \otimes B')^\top \)

**Remark**

6. Verify (iv) \((A \otimes \frac{1}{2})\)
Matrix Normal Distributions

Suppose $X$ obtained from $X = (x_{1}, x_{2}, \ldots, x_{n})^{T}$

has a joint normal distribution with

\[
E_{\mu} \mu_{j} \mu_{i} = \mu_{j} \mu_{i} \quad \text{and} \quad \text{cov}(x_{j}, x_{i}) = \Delta \sigma_{j} \sigma_{i} \quad \Rightarrow \quad \text{cov}(X) = \Delta \otimes \Sigma
\]

for $j, i \in 1:n$, $i, j \in 1:p$.

Then we write

\[
\begin{bmatrix} x_{1} \cdots x_{n} \end{bmatrix} \sim \mathcal{N}_{p \times n}(\mu, \Sigma \otimes \Delta)
\]

\[
\mu = (\mu_{1}, \mu_{2}, \ldots, \mu_{n})
\]

\[
\text{cov}(x_{j}, x_{i}) = \Delta \sigma_{j} \sigma_{i} \quad \Rightarrow \quad \text{cov}(\nu_{j}, \nu_{i}) = \sigma_{j} \sigma_{i} \Delta
\]

- Normal Covariances

> Not all multivariate normal distributions for $X$

\[
\text{Fact:}
\]

(i) $\mathcal{N}_{p}(\mu, \Sigma) \Leftrightarrow X \sim \mathcal{N}_{p \times n}(\mu, \Delta \otimes \Sigma)$

(ii) $X \sim \mathcal{N}_{n}(\mu, \Delta \otimes \Sigma)$
(iii) If $Z$ and $\Delta$ have spectral representations

\[ Z = W D W^T \quad \text{and} \quad \Delta = W D^2 W^T \]

\[ p \times k \times k \times k \times p \quad 2 \times k \times k \times k \times k \times n \]

then

\[ Z = \mu + W D^{\frac{1}{2}} Z D^{\frac{1}{2}} W \]

\[ Z = \mu + Z D^{\frac{1}{2}} \]

where $Z$ is the "standard normal matrix"

\[ Z \sim N_{k \times k} \left( 0, I_k \otimes I_k \right) \]

\[ \left[ z \sim N(0, I) \right] \]

**Theorem 1**

\[ X \sim N_{p \times n} \left( \mu, \Sigma \otimes \Delta \right) \quad \text{and} \quad Y = AXB \]

\[ \text{Imp} \]

\[ Y \sim N_{n \times b} \left( A\mu B, A\Sigma A' \otimes B\Delta B \right) \]
We observe \( x_i \sim \mathcal{N}(\mu, \Sigma) \), \( i = 1, 2, \ldots, n \), with \( \Sigma \) known
d to have diagonal elements \( \Sigma_{ii} = \sigma_i^2 \) and compute
\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (i = 1, 2, \ldots, p)
\]
to test the null hypothesis \( \mu = 0 \). What is the correlation matrix of \( \bar{\mathbf{z}} = \left( \begin{array}{c} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{array} \right) \)?

**Rotational Invariance** Write \( \mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}) \) as
\[
\mathbf{X} = \left( \begin{array}{c} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{array} \right).
\]
The individual row vectors \( \mathbf{v}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}) \)
individually are spherically symmetric: \( \mathbf{v} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}) \).

\[
\text{Hence show that } \mathbf{v}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}).
\]
It turns out $v_1, v_2, \ldots, v_p$ together also have spherical symmetry, even though they are correlated; \[ \text{corr}(v_i, v_j) = \mathbb{I}_{i \neq j} \]

**Corollary** For any non-orthogonal matrix $\Gamma$,

\[ \Gamma (v_1, v_2, \ldots, v_p) \sim (v_1, v_2, \ldots, v_p) \quad \text{[or $X' \sim X$]}. \]

**Proof** \[ X' \sim N_{p, \mathcal{C}} (\mu, \Sigma) \Rightarrow X' \sim N_{p, \mathcal{C}} (0, \Sigma) \]

An equivalent statement is that the $p$-dimensional subspace of $\mathbb{R}^n$, $X$, is uniformly distributed over all possible orientations ("Harm measure").

(Though they are inter-correlated.)
Removing Means

Let \( \mathbf{P} = I - \mathbf{x} \mathbf{x}^T / n \), then the projection matrix orthogonal to \( \mathbf{x} \) is

\[
\mathbf{Y} = \mathbf{X} \mathbf{P} = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}), \quad \bar{x} = \frac{\sum x_i}{n},
\]

They

(i) \( \mathbf{Y} \sim \mathcal{N}_n \left( \mathbf{0}, \frac{1}{n} \mathbf{P} \right) \)

\[
\mathbb{E} \mathbf{Y} = \mathbf{0}, \quad \mathbb{V} \mathbf{Y} = \frac{1}{n} \mathbf{P}
\]

(ii) \( \mathbf{Y} \perp \bar{x} \sim \mathcal{N} \left( \mu, \frac{1}{n} \mathbf{P} \right) \)

\[
\mathbb{E} (\bar{x} | \mathbf{Y}) = \frac{\mathbf{P} \mathbf{Y}}{\mathbf{P} \mathbf{P}^T} \quad \text{and} \quad \mathbb{V} (\bar{x} | \mathbf{Y}) = \frac{1}{n} \mathbf{P}
\]

(iii) \( \Rightarrow \mathbf{S} = \mathbf{Y} \mathbf{Y}^T = \left( \frac{1}{n} \sum (x_i - \bar{x})(x_j - \bar{x}) \right) \mathbf{I} \mathbf{Y} \)

(iv) Suppose \( \mathbf{\Gamma} = \left( \Gamma_1, 1/n \right) \), then

\[
\mathbf{Z} = \mathbf{X} \mathbf{\Gamma},
\]
Then
$$\mathbb{E} \sim \mathcal{N}_{p \times n} (0, \Sigma \otimes I),$$

i.e.,
$$\mathbb{E} \sim (\mathbb{E}_1, \ldots, \mathbb{E}_n) \text{ where } \mathbb{E}_i \sim \mathcal{N} (0, \Sigma).$$

But
$$\mathbb{E} = X \Gamma = \frac{1}{\sqrt{n}} \mathbb{E}^T, \quad (\text{since } \mathbb{E}^T = \Gamma),$$

$$= Y \Gamma,$$

so $\mathbb{E}$ is just $Y$ expressed in an $n \times 1$-dimensional coordinate system in $\mathbb{R}^n$.

Hence $X \sim \mathcal{N} (\mu, \Sigma \otimes I)$ and $Y = \frac{1}{\sqrt{n}} X^T$.

$$\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{n}}, \quad \Sigma = I - \frac{1}{\sqrt{n}} I \otimes I.$$ Show this.

\[ \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{E}_{ij} = 0 \quad \text{for all } i \text{ and } j. \]
\( \mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \Delta) \), where

\[ \mathbf{\delta}_i^y = \mathbf{\delta}_i - \mathbf{\delta}_i - \mathbf{\delta}_i - \mathbf{\delta}_i, \quad i, \delta_i = 1, \ldots, p \]

and

\[ \Delta_{ij}^y = \Delta_{ij} - \Delta_{ij} - \Delta_{ij} + \Delta_{ij}, \quad i, j, \gamma = 1, \ldots, r \]

"dot" indicating an average over the missing index.

\[ \sum_i \mathbf{\delta}_i^y = \sum_i \Delta_{ij}^y = 0 \quad \text{for all } i \text{ and } j. \]

Note: Denoising usually reduces correlations.

Example: "prostate" \( p = 6033 \) genes \( n = 102 = 50 + 52 \) men.

50 controls, 52 prostate cancer patients.

\( \mathbf{X} = 6033 \times 50 \) matrix of expression values for the
controls: microarrays standardized to have mean 0 and 1.

* The \( \binom{50}{2} = 1225 \) column-wise correlations \( \sim (0.0592, 0.0234^2) \)

* With gene row means subtracted \( \sim (-0.0204, 0.0249^2) \)

**Handout 10**

(a) Why do the demeaned correlations have \( \overline{\text{cor}} = -0.0234 \)?

(b) Examine the 50x50 correlation matrix for the original (undemeaned) data. Where do the large values—say those \( \geq 0.125 \)—come from?
figB1. All 1225 column correlations from the 6033x50 matrix of controls, prostate data

demeaned data
emcor~(-.0204,.0247^2)

original data
cemcor~(.0592,.0234^2)
Spherical Symmetry and the t-test. The spherically invariant distribution of the multivariate normal distribution plays a crucial role in classical hypothesis testing. Students one-sample t-test offers the simplest example:

\[ x_i \sim \mathcal{N}(\mu, \sigma^2) \quad i = 1, 2, \ldots \quad H_0: \mu = 0 \]

\[ t = \frac{\hat{\mu} - \mu_0}{\|n - 1\|^{-\frac{1}{2}}} \quad \|n - 1\|^{-\frac{1}{2}} \quad \text{sign} \cdot \tan A \]

- One-sided t-test rejects for \( x/n \) inside a spherical cap of radius \( A_{0.05} \) centered at \( 1/n \) on the unit sphere, where the cap contains 5\% of the area of \( S_1 \).

Homework II: Compute \( A_{0.05} \) for \( n = 10, 20, 40, 80 \).
The p-value for $x$ is the area of the spherical cap through $x$ through $x_0$, centered at the "north pole" $2/\sqrt{n}$, divided by the area of the whole whole sphere.

The 2-sample $t$-test is essentially the same, now with $x$ projected into $\mathbb{R}^p$, and the north pole now

$$d = \left( \frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_p}, \frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_2} \right).$$

postulate gives 6933 two-sample t-values, one for
figB2. Two-sample t-values for the 6033 genes in prostdata; Benjamini-Hochberg(.1) test gives 28 'sig' genes, t>3.43
each gene comparing the control and cancer groups. How can we decide which genes are genuinely different in the two groups.

Benjamini-Hochberg procedure: Order the $p$-values

\[ p_{(1)} < p_{(2)} < p_{(3)} \ldots < p_{(n)} \leq \frac{q}{n} \]

"most significant."

"Declare "significant" those cases with $p_{(i)} \leq p^*$ where

\[ p^* = \max \left\{ \frac{q}{n} \cdot \frac{k}{N} : k = 1, \ldots, n \right\} \]

for some precision constant $q \in (0, 1)$.\]
BH Theorem Expected proportion of "false discoveries" ≤ q.

For postdata with q = .1. The 28 genes with t-value > 3.53

(p-value < .000438) "significant".

If we could look at the 6,033 vectors \( \mathbf{a} \) obtained
by projecting \( \mathbf{x} \) orthogonally to \( \mathbf{z} \), we'd see that there
were too many near the north pole \( \mathbf{a} \). As a matter
of fact, the cap just containing these significant 28 is overpopulated
by a factor > 10, compared to its 100-dimensional area.
B2. Normal Matrices, Triangular Coordinating, and Hui Lemma

\[ X = \begin{pmatrix} -v_1^* \\ -v_2^* \end{pmatrix} \sim N_p (0, \theta I) \]

Can be thought of as composed of two parts:

\[ L = \frac{1}{2} \sum_{x \in \mathcal{L}} (x) = \sum_{e \in \mathcal{L}} \langle v, v \rangle, \]

and the lengths and angles between the vectors \( v_1, \ldots, v_p \) in \( L \).

Corollary 1 of Section B1 says that \( L \) is uniformly distributed over the space of \( p \)-dimensional linear subspaces of \( \mathbb{R}^n \), say

\[ L \sim \text{Uniform}(\mathcal{L}) \quad [= \mathcal{U}(\mathcal{L})] \]
The joint distribution of the lengths and angles of \( y_1, \ldots, y_p \) within \( L \) is the **Wishart Distribution**. The Gram-Schmidt triangular representation leads to an explicit density function for the Wishart.

**Jacobian to Triangular Coordinates** Let \( V = X' = (y_1, y_2, \ldots, y_p) \),

and as in A2.8, write \( V \) in Gram-Schmidt form

\[
V = W' T
\]

The inner product matrix

\[
S = XX' = V V' = T' W' W T = T' T
\]
then contains all the information about the lengths and angles.

\[ \text{Lemmas 1} \quad \text{The integral Jacobian from } \mathbf{x} \rightarrow \mathbf{T} \quad \text{is} \]

\[ J(\mathbf{x} \rightarrow \mathbf{T}) = \epsilon_i \prod_{y \in \mathcal{Y}} t_{y}^{-\frac{1}{2}} \]

where \( \epsilon_i = \frac{1}{2} \pi^{\frac{2}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{2}} \)

See page 2.5

Proof: Consider \( \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \), \( \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \).

\[ \forall \mathbf{x} \ni \mathbf{x} \sim \mathcal{N}(\mathbf{0}, 1) \]

\[ f(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2} \mathbf{x}^T \mathbf{X} \mathbf{x}} = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2} \text{tr} \mathbf{X} \mathbf{X}'} = e^{-\frac{1}{2} \text{tr} \mathbf{X} \mathbf{X}'} / (2\pi)^{-\frac{d}{2}} \]

\[ = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2} \text{tr} \mathbf{T} \mathbf{T}'} = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2} \sum_{y \in \mathcal{Y}} t_{y}^2} \]
Now \( \nu = (x_1, x_2, \ldots, x_n) \sim \mathcal{N}(0, I_n) \) so \( \nu \) has a random direction in \( \mathbb{R}^n \); since by construction

\[
\nu_1 = t_1 \cdot w_1,
\]

\[
\frac{\nu_1^2}{n} \sim \chi^2_n.
\]

\[
\nu_2 = (\nu_1, \nu_2, \ldots, \nu_n) \sim \mathcal{N}(0, I_n), \text{ independent of } \nu_1,
\]

\[
\nu = \frac{\nu_1}{\sqrt{n}} + \frac{\nu_2}{\sqrt{2}} = \frac{\nu_1}{\sqrt{2}} + \frac{\nu_2}{\sqrt{2}},
\]

the decomposition of \( \nu_1 \) into \( \mathbb{R}^1 \) and \( \mathbb{R}^{n-1} \). Because of independence, we can think of \( \nu_1 \) as a fixed direction in \( \mathbb{R}^n \), so \( \frac{\nu_1}{\sqrt{n}} \), the projection of \( \nu_2 \) along \( w_1 \), is \( \mathcal{N}(0, 1) \)

\[
\frac{\nu_1}{\sqrt{n}} \sim \mathcal{N}(0, 1).
\]
Also, \[ t_{2n} = \| U \|_2^2 \sim \chi^2. \]

Continuing in this way, we see that

\[
\begin{align*}
t_{i} & \sim \chi^2_{(0,1)}, \\
t_{i+1} & \sim \chi^2_{(1,1)}, \\
& \vdots \\
& \sim \chi^2_{(i,1)}
\end{align*}
\]

\( i = 1, 2, \ldots \)

Giving

\[
T(t) = \frac{\prod_{i=1}^{n} e^{-\frac{t_i^2}{2}}}{\prod_{i=1}^{n} \Gamma\left(\frac{\alpha_i + \beta_i}{2}\right) \frac{1}{\Gamma(\frac{\alpha_i}{2})} \frac{1}{\Gamma(\frac{\beta_i}{2})}}
\]

\( i \in \text{comb. x dist.} \)

Finally, by the trick at the end of Section A3,

\[
\overline{T} (x \rightarrow t) = \frac{1}{\overline{T}(t)} \int dx^2
\]
Useful Device (We've just used):

Suppose we can find a simple situation when

\[ f(x) = g(t) \]  \text{ density only depends on } g(t)

is a known function of \( t \)

and we know \( f^T(t) \). Then \( f^T(t) = g(t) \int (x \rightarrow t) \),

we get

\[ J(x \rightarrow t) = f^T(t) / g(t). \]

In this case of Lemma 1, the simple situation is

\[ X \sim N_p (0, I \otimes I) \cap \mathbb{R}^n \sim N(0, I). \]

This is easy to write down \( g(t) \). The calculation proceeds by finding \( f^T(t) \) iteratively: \( t, \frac{n}{2}, \ldots \). etc.
Handout 2 \#41 Carry out step 3 of "continuing in this way.

(c) Explicitly verify the value of the constant $c$.

Note: The random orthogonal matrix $W$ is independent of $T$. Another way to say the lemma is

$$ J(x \to T^p L) = \int_{T_x = P_x} \frac{p^T u_0 (L)}{\lambda_x} dx $$

where $u_0 (L)$ is the invariant density on $p$-dimensional subspaces of $R^n$. This follows from

$$ J(x \to T^p L) = \frac{f^{T^p L}(x, L)}{f^{T_p}(x)} \frac{f^L(x, e)}{f^L(x)} $$

$$ = J(x \to T) u_0 (L). $$
Counting coordinates. It takes

\[
\frac{(n-1) + (n-2) + \cdots + (n-p)}{w_1 - w_2 - \cdots - w_p} \quad \Theta = np - \frac{p(p+1)}{2}
\]

coorindates to specify \( L \). The triangular matrix \( T \) has \( p(p+1)/2 \) free coordinates, so \( (L,T) \) together have \( np \)

dree coordinates, the same as \( X \). Therefore \( T(X \rightarrow \tilde{X}) \)

is a regular, not integer, Jacobian, the mapping being 1-1.

**Corollary** If \( X \sim \mathcal{N}_{p \times n}(0, \Sigma I_n) \), with \( \Sigma > 0 \), then

\[
f_T(t) = \frac{2t^{(p/2)-1}}{\Gamma(p/2)} \frac{p}{\pi^{p/2}} e^{-\frac{1}{2}tr \Sigma t}
\]

\( (p = t' \Sigma t) \).

**Homework 2** Verify \( T \).
Mapping $T \rightarrow S$ The mapping $T \rightarrow S = T^T$ takes the triangular matrix $T$ into the symmetric matrix $S$, both of which have $p(p+1)/2$ free coordinates.

Lemma 2 \[ J(T \rightarrow S) = \left[ \begin{array}{cc} \frac{p}{2} & \frac{p+1}{2} \\ \frac{p+1}{2} & \frac{p}{2} \end{array} \right] \]


Examine the $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ matrix (\begin{array}{ccc} \text{exchange} & \text{exchange} \\ \text{etc.} & \text{etc.} \end{array}) and notice it is of form $\left( \begin{array}{cc} A & 0 \\ 0 & B \end{array} \right)$.

Hence Lemma. The mapping $X \rightarrow S = XX^T$ has integer Jacobian

\[ J(X \rightarrow S) = \frac{1}{2}(\frac{p}{2} - \frac{p+1}{2}) S_1 \]

where

\[ \frac{1}{2} \left( \frac{\frac{p}{2} - \frac{p+1}{2}}{2} \right) \]
Proof \[ J(X \to S) = J(X \to T) J(T \to S) \]

Hark 4: Complete the proof.

Notice we can also write Hsieh Lemma as

\[ J(X \to S) = \frac{1}{n} [V_0(v_1, v_2, \ldots, v_n)]^{n-p-1} \]

Since \( |S| = V_0^3 \).

Hark 5: Vector \( v_1, v_2, \ldots, v_n \) are chosen independently and uniformly within the unit sphere in \( \mathbb{R}^n \). For positive integer \( m \), calculate \( E[|V_0(v_1, v_2, \ldots, v_n)|^{2m}] \).

* The next section uses Hsieh Lemma to prove the Wishart distribution for \( S \).
B3. The Wishart Distribution

If \( X \sim \mathcal{N}_p \left( \mathbf{0}, \Sigma \otimes I \right) \), then \( S = X'X = \sum_{i=1}^n x_i x_i' \) is said to have the Wishart distribution,

\[ S \sim \mathcal{W} \left( \Sigma, n, p \right). \]

Fisher derived distribution for case \( p = 2 \), then Wishart (1928) gave the general result \( S \).

\[ p = 1: S = \frac{\Sigma x^2}{\mathcal{N} \left( 2 \sigma^2 \right)}, \text{ so } S \sim \sigma^2 x^2. \]

**Theorem 2** If \( \Sigma > 0 \) then \( S > 0 \) with probability 1, and has density (generalization of central \( x^2 \))

\[
f \left( s \right) = c_3 \frac{n-1}{2} e^{-\frac{1}{2} \left( \frac{s}{n-2} \right)} / \left( 2 \right)^{n/2} \]

\[ (s > 0) \]

with

\[ c_3 = \left[ 2 \frac{n}{n-2} \right]^{\left( n-1 \right)/2} \left( \pi \right)^{n/2}. \]
Proof: \[ f^X(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}} = (2\pi)^{-\frac{n}{2}} \left| \Sigma' \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{x}' \Sigma'^{-1} \mathbf{x}} = (2\pi)^{-\frac{n}{2}} \left| \Sigma' \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{x}' \Sigma'^{-1} \mathbf{x}} \]

Thus, using Hsu's Lemma,

\[ f^X(x) = f^{x}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \]

Check: For \( f^X(x) \), we get

\[ f^X(x) = \frac{a^2 e^{-\frac{1}{2} \frac{x^2}{a^2}}}{\sqrt{2\pi}} \quad (a > 0) \]

Homework: Why is this correct?

Properties and Facts: The density function \( f^X(x) \) is not much used - instead, various transformation properties are invoked.
(i) \( S \sim \mathcal{W}(\xi; \eta, \theta) \) and \( \mathcal{W}(A^T \xi A; \eta, \theta) \)

\[ \Rightarrow \tilde{S} \sim \mathcal{W}(A \hat{x} A^T; \eta, \theta) \quad \text{[} \tilde{S} = \sum_{i=2}^{n} \xi_i, \tilde{\xi} = \sum_{i=2}^{n} \xi_i \text{ where} \]

\[ \tilde{\xi} = A \tilde{x}, \text{ where} \]

\[ H_{\text{null}}: \text{Suppose } A \text{ non-singular. Show that the mapping} \]

\[ S \rightarrow \tilde{S} = A \hat{x} A^T, \text{ has } \text{Jacobian } J(S \rightarrow \tilde{S}) = (|A|)^{-\frac{p+1}{2}} \]

[Hint: use (i).]

(ii) \( S_1 \sim \mathcal{W}(\xi; \eta_1, \theta) \) \( \parallel \) \( S_2 \sim \mathcal{W}(\xi; \eta_2, \theta) \) \( \Rightarrow \) \( \tilde{S}_1 \sim \mathcal{W}(\xi; \eta_1, \theta) \) \( \Rightarrow \) \( \tilde{S}_2 \sim \mathcal{W}(\xi; \eta_2, \theta) \)

\[ \sum_{i=2}^{n} \tilde{x}_i = \sum_{i=2}^{n} \xi_i + \sum_{i=2}^{n} x_i. \]

(iii) \( \text{Suppose} \)
Matrix of Means

\[ X \sim N \left( \mu, \Sigma \otimes I \right) \text{ and } \Gamma \text{ such that } \eta \Gamma = 0 \]

Then \( Y = \Gamma X \) has

\[ YY' \sim W \left( \mathbb{R}^p ; m, \rho \right) \]

\[ p \times p \]

\[ Y \sim N \left( \Lambda \eta \Gamma \mu, \eta \Gamma \Sigma \Gamma' \eta \right) \text{ if } \eta \in \mathbb{R}^p \]

\[ \Lambda \text{ is the projection matrix into } \mathbb{R}^p \text{ with } \Lambda \eta = 0 \]

\( Y \sim W \left( \mathbb{R}^p ; m, \rho \right) \).

\[ YY' = X \eta \Gamma \mu \eta \Gamma' X' = X \eta \Gamma \eta \Gamma' X', \text{ so result follows} \]

From (iii) \( X \eta \Gamma \eta \Gamma' \)

\[ \sim X \Gamma \eta \eta \Gamma' \]
Removing the means \[ \text{Let } P = I - \bar{x} \bar{x}' / n, \]

so \[ \bar{x} \sim N \left( \mu, \sigma^2 I \right) \]

\[ Y = X P = (x - \bar{x}) P = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}) \]

where \[ \bar{x} = \frac{\sum x_i}{n}. \]

Then all give

\[ \sum_{i=1}^{n} (x_i - \bar{x}) (x_i - \bar{x})' \sim W \left( \Sigma; n-1, \rho \right) \]

**Homework 3:** Let \[ \bar{x} = Q X P \] when \( Q = I - 11' / p \), \( P = I - \bar{x} \bar{x}' / n \).

Show that \[ \bar{x} \sim W \left( \bar{x}; n-1, \rho \right) \] when the elements of \( \bar{x} \)

\[ \bar{x} \text{ are } \bar{x} = \frac{\sum x_i}{n}, \bar{x} = \frac{\sum x_i}{n}, \bar{x} = \frac{\sum x_i}{n}, \ldots, \bar{x} = \frac{\sum x_i}{n}. \]

\[ \left( \sum_{i=1}^{n} \bar{x}_i = 0 \right) \text{ for all } \bar{x}_i. \]
Hankel 4 / For a positive integer, find $E(\sum_{i=1}^{n^2} z_{ii}^{2n})$.

(Note that gives $E(V_0^{2n})$ for $X = (x_{ij})$, $V_0 = \text{Variance}(x_{11})$.)

(v) Decomposition Property Suppose we observe

\[
\begin{pmatrix}
    x_1^a \\
    x_2^a \\
    \vdots \\
    x_a^a
\end{pmatrix}
\sim \mathcal{N}
\begin{pmatrix}
    0 \\
    0 \\
    \vdots \\
    0
\end{pmatrix}
\begin{pmatrix}
    S_{xx} & S_{xy} \\
    S_{yx} & S_{yy}
\end{pmatrix}
\end{pmatrix}
\begin{pmatrix}
    \theta_1 \\
    \theta_2 \\
    \vdots \\
    \theta_{a-1}
\end{pmatrix} + \mathcal{N}(0, \Sigma),
\]

\[
\Sigma =
\begin{pmatrix}
    S_{xx} & S_{xy} \\
    S_{yx} & S_{yy}
\end{pmatrix}.
\]

Giving data matrix

\[
\begin{pmatrix}
    x_1^a \\
    x_2^a \\
    \vdots \\
    x_a^a
\end{pmatrix}
\sim \mathcal{N}
\begin{pmatrix}
    0 \\
    0 \\
    \vdots \\
    0
\end{pmatrix}
\begin{pmatrix}
    \Sigma_{xx} & \Sigma_{xy} \\
    \Sigma_{yx} & \Sigma_{yy}
\end{pmatrix}
\begin{pmatrix}
    \theta_1 \\
    \theta_2 \\
    \vdots \\
    \theta_{a-1}
\end{pmatrix} + \mathcal{N}(0, \Sigma),
\]

and joint Wishart matrix \[
\begin{pmatrix}
    \Sigma_{xx} & \Sigma_{xy} \\
    \Sigma_{yx} & \Sigma_{yy}
\end{pmatrix} \sim \mathcal{W}(\Sigma, \eta, \rho^2).
Define

\[ \hat{P} = X (X' X)^{-1} X' \]

Projector Matrix into \( \mathbb{R}^p \( X \) \)

\[ \hat{P}_x = \hat{P} \]

Projection Matrix into \( \mathbb{R}^p \( X \) \)

Both \( \hat{P} \) and \( \hat{P}_x \) are random, depending on \( X \). We can

(randomly) decompose \( Y \) into

\[ Y = \hat{Y} + \hat{Y}' \]

\[ \sim \hat{Y} \sim_X \hat{Y}' \sim \]

and likewise decompose \( \Sigma_y = \hat{Y} \Sigma_y \hat{Y}' \)

\[ \Sigma_y = (\hat{Y} + \hat{Y}) (\hat{Y} + \hat{Y})' \]

\[ = \hat{Y} \hat{Y}' + \hat{Y} \hat{Y}' \]

\[ = \Sigma_y \Sigma_y \Sigma_y \Sigma_y \]
Giving the familiar Grammian decomposition

\[ S_y = y_y = S_y - S_x S_{xx}^{-1} S_x y \]

as in Section A2, except that \( y \) is normally defined.

**Decomposition Lemma**

\( S_{xx} \sim W(\Sigma_x; n, p) \) \[ Claim \]

\[ S_x \sim \Sigma_x \begin{bmatrix} \Sigma_x^{-1} & 0 \\ 0 & \Sigma_y^{-1} \end{bmatrix} S_x \Sigma_y^{-1} \]

where

\[ \Sigma_y^{-1} = \Sigma_y - S_{xy} S_{xx}^{-1} S_{yx} \]

\( S_{xy} \sim W(\Sigma_{xy}; n, p, q) \)

\( S_{yx} \sim W(\Sigma_{yx}; n, p, q) \)

(b) \( S_{yx} \) and \( S_{yy} \), the two parts of \( S_{yy} \), are conditionally independent given \( S_{xy} \). \[ \text{[Implication:} S_{yy} \parallel S_{xy}, \text{since the conditional distribution does not depend on } S_{xy} \text{]}. \]
Proof: \[
Y \sim N \left( \mathbf{X} \mathbf{X}^\top \mathbf{\Sigma}_x, \mathbf{\Sigma}_x \otimes I \right) \quad \text{for } i = 1, \ldots, N,
\]
This gives
\[
\begin{align*}
\hat{Y} &= Y \mathbf{P} \sim N \left( \mathbf{X} \mathbf{X}^\top \mathbf{\Sigma}_x, \mathbf{\Sigma}_x \otimes \mathbf{P} \right),
\end{align*}
\]
\[
\mathbf{S} = \mathbf{X} \mathbf{X}'
\]
Implying,
\[
\hat{Y} \mathbf{P} \sim N \left( \mathbf{X} \mathbf{X}^\top \mathbf{\Sigma}_x, \mathbf{\Sigma}_x \otimes \mathbf{P} \right),
\]
Because this only depends on \( \mathbf{X} \) through \( \mathbf{S}_{xx} \), we get (6),
\[
\begin{align*}
\mathbf{S} \mathbf{X} \sim N \left( \mathbf{X} \mathbf{X}^\top \mathbf{\Sigma}_x, \mathbf{\Sigma}_x \otimes \mathbf{S}_{xx} \right)
\end{align*}
\]
In other words, the gap matrix of inner products between \( Y \) and \( \mathbf{X} \).
has a Kronecker type of matrix normal distribution given $\mathbf{S}_{xx}$.

Likewise for (c): $\mathbf{Y} = \mathbf{Y}^*_{x \sim \mathbf{F}}$ and $\mathbf{F}$ give

$$
\mathbf{Y} \mid \mathbf{X} \sim \mathcal{N}
\left(
\begin{bmatrix}
\mathbf{X} \\
\mathbf{X} \\
\mathbf{X}
\end{bmatrix},
\begin{bmatrix}
\mathbf{P} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{P} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{P}
\end{bmatrix}
\right)
$$

so by property (iv) on B34,

$$
\mathbf{Y} \mid \mathbf{X} \sim \mathcal{N}
\left(
\begin{bmatrix}
\mathbf{X} \\
\mathbf{X} \\
\mathbf{X}
\end{bmatrix},
\begin{bmatrix}
\mathbf{P} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{P} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{P}
\end{bmatrix}
\right).
$$

By property 5, verify (d), that $\mathbf{S}_{xx}$ and $\mathbf{S}_{yy}$ are conditionally independent given $\mathbf{S}_{x \sim \mathbf{F}}$. (Hint: First show that $\mathbf{Y} \mid \mathbf{Y} \mathbf{X}) \sim$)

The case $p = q = 1$
\( \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_1 \\ x_n \end{pmatrix} \sim \begin{pmatrix} \frac{x^2}{\sigma_{xx}} \end{pmatrix} \sim \mathcal{N}_2(\bar{\mathbf{x}}, \begin{pmatrix} \sigma_{xx} \sigma_{xy} \\ \sigma_{xy} \sigma_{yy} \end{pmatrix}) \)

(1) \( S_{xx} = \|x\|^2 \sim W(\sigma_{xx}, n, 1) = \sigma_{xx}^2 \)

(2) \( S_{yy} = y \cdot x \|x\|^2 \sim W\left(\frac{\sigma_{yy} \|x\|^2}{\sigma_{xx}}, \frac{\sigma_{xy} \|x\|^2}{\sigma_{xx}} \right) \)

where \( \sigma_{xy} = \sigma_{xy}^2 / \sigma_{xx} \)

(3) \( \frac{1}{\sigma_{yy}^2} = \|y^\parallel^2 / \|x\|^2 \)

(4) \( y \cdot x \|y^\parallel^2 / \|x\|^2 \)

Function of a Wishart matrix

Inverse \( \check{S} = S \) where \( S \sim W(\Sigma; n, q), n > 0. \)
Lemma The mapping \( \tilde{S} = \tilde{S}' \) of the space of symmetric positive definite \( p \times p \) matrices into itself has Jacobian

\[
J(S \rightarrow \tilde{S}) = \tilde{S}^{-(p+1)} = S^{p+1}
\]

Proof. Makes use of a convenient trick. Suppose we change \( S \) to \( S + dS \), \( dS \) symmetric infinitesimal matrix. Notice that

\[
\left[ S + dS \right]\left[ S' - S'(dS) S' \right]^2 S' = I + \left( dS \right) S' - \left[ (dS) S' - (dS) S' \right]^2
\]

\[
= I + O(dS^2),
\]

or, letting \( T = S' \),

\[
dT = -T(dS)T
\]
This means that locally, near some given value of $s$, the mapping $S \rightarrow T$ behaves like the linear transformation $dT = -T(ds \mid T)$. Homework 2 on B3.3.

(With $A$ equ. $T$) gives

$$J(S \rightarrow T) = J(\|ds \rightarrow dT\|) = \lambda A$$

$$= \lambda f^{-2p+1} = J5f^{-p+1}$$

**Corollary**: If $S \sim W(\frac{\theta}{\lambda}, \lambda, \rho)$, $\rho > 0$, then $T = S'$

**Density**

$$f_T(t^r) = c_3 \left(\frac{\lambda t^{2p+1}}{2}\right) e^{-\frac{1}{2} \lambda \cdot \frac{\theta}{\lambda} r^2} / |S|^2$$
with $c_3 = \sum_{2 \leq s \leq \frac{p}{2} \frac{p-1}{2}} \prod_{j \in s} \frac{\Gamma \left( \frac{n-1}{2} \right)}{\Gamma \left( \frac{n-2s+1}{2} \right)}$.

Hence verify ↑

The Correlation Matrix

Let $D = \left( \begin{array}{c} d_i \\ d_p \end{array} \right)$, $d_i = \sqrt{\lambda_i}$,

and

$R = D^T S D$.

so $R$ has diagonal elements 1 and off-diagonals

$r_{ij} = \frac{d_i d_j \cdot \sum d_i d_j x_{ij}}{\sqrt{d_i^2 \cdot d_j^2}}$

If $S = \sum_{i \neq j} (x_{i} - \bar{x}_i)(x_{j} - \bar{x}_j)$, then $r_{ij} = \text{sample correlation coeff between variables } i \text{ and } j$. 

$\sum_{i \neq j}$ and $j$. 
Lemma \( J(S \rightarrow D, R) = 2^p |D|^p \)

Hence verify (by direct evaluation)

Condition If \( S \sim W(I; \nu, p) \) then \( D \) and \( R \) independent with

\[
\mathbb{V} \mathbb{E} (\mathbb{R}) = \mathbb{E} (\mathbb{R})^{\frac{n-2}{2}} \quad \text{and} \quad \mathbb{V} \mathbb{E} (\mathbb{\mathbb{X}}) = \mathbb{E} (\mathbb{\mathbb{X}})^{\frac{2}{2}}
\]

Proof \( f(\mathbb{R}) = f(S) J(S \rightarrow D, R) \)

\[
= \mathbb{E} (|D|^p |S|) \quad \mathbb{E} (|D|^p |S|) = 2^p |D|^p \quad e^{-\frac{1}{2} \|S\|_2^2}
\]

But \( |S| = |D|^p |R| \) and \( \text{tr} S = \text{tr} D R = \text{tr} R = \sum_{i,j} R_{ij} = \sum d_{ij}^2 \)

Homework Finish the proof. What "goes wrong" if \( S \neq I_p \)?
(vi) Moments of the Wishart

\[ x \sim N_p (\Sigma, \Sigma) \rightarrow X = (x_1, x_2, \ldots, x_p) \rightarrow S = X'X \]

\[ \hat{S} = X'X / n \sim W_p (\Sigma, n, p) / \Sigma \]

[If \( x \sim N_p (\Sigma) \) then \( \hat{S} = \sum (x_i - \bar{x})(x_i - \bar{x}) / (n - 1) \sim W_p (\Sigma, n - 1, p) / (n - 1) \)]

**Theorem**

\[ \hat{S} \sim \left( \begin{array}{c}
\frac{1}{n-p-1} \\
\frac{1}{n-p-1} \\
\end{array} \right) \left( \begin{array}{c}
\frac{1}{n-p-1} \\
\frac{1}{n-p-1} \\
\end{array} \right) \]

\[ \text{Mean Covariance} \]

\[ \hat{S} \]

\[ \left( \begin{array}{c}
\frac{1}{n-p-1} \\
\frac{1}{n-p-1} \\
\end{array} \right) \left( \begin{array}{c}
\frac{1}{n-p-1} \\
\frac{1}{n-p-1} \\
\end{array} \right) \]

where

\[ \Gamma = \left( \begin{array}{cc}
\frac{1}{n-p-1} & \frac{1}{n-p-1} \\
\frac{1}{n-p-1} & \frac{1}{n-p-1} \\
\end{array} \right) \]

\[ \text{Proof from property (viii), B)2.} \]

In other words, the symmetric p×p matrix \( \hat{S} \) is unbiased for \( S \); it has \( p(p+1)/2 \) free elements whose covariance are \( \Gamma \).

\[ \text{Hint:} \frac{\mathbf{A}}{\mathbf{S}} / \mathbf{F} \text{ p=2 and } \text{cor}(x_{11}, x_{22})=0, \text{ show that cor}(\hat{S}_{11}, \hat{S}_{22})=\sigma^2 \]
Suppose $X = (x_1, \ldots, x_n) \sim \mathcal{N}(0, \Sigma_{X})$.

When $y_i = 1$ for $i = 1, \ldots, n$, so $X \sim (0, \Sigma_{X})$.

Let $\hat{\Sigma} = \hat{X}' \hat{X} / n$.

Lemma: $\hat{\Sigma} \sim (\Sigma_{X}, \Gamma / n_{\text{eff}})$ where the effective sample size $n_{\text{eff}}$ is

$$n_{\text{eff}} = 1 / \left[ \frac{1}{n} + \frac{\alpha^2}{n} \right]$$

with $\alpha = \left[ \sum_{y_i = 1} g_{yy} / (\Gamma) \right]^{1/2}$.

($\Gamma$ as in Theorem).

In other words, $\hat{\Sigma}$ is still unbiased for $\Sigma_{X}$ but

with increased variability if the columns are inter-correlated.
A Microarray Example \( p \leq 44 \) people measured on \( n = 20426 \) genes. Data matrix \( X = (x_{ij}; x_{1i}, x_{2i}, \ldots, x_{pi}) \) has been centered \( \mu \) and row scaled so that the means and standard deviations for each column and each row are \((0, 1)\).

- \( \tilde{x} = \text{corrected version of } \sqrt[2]{\frac{\sum_{j \in y_j} (\tilde{x}_{ij} - \tilde{\mu}_j)^2}{(n_j - 2)}} \) 
  \( \approx 204 \text{ Million} \)

\( r_{eff} = 17.2 \)

- Most of the standard microarray analysis methods assume that the rows of \( X \) \((v_i)\) are simple independently.

In this case I had some doubts.
figB3. First eigenvector of SigmaHat matrix from 20426x44 microarray matrix: e[1]/sum(e[i])=.207

Coordinate 32 vs coordinate 31, sample correlation = .805
figB4. 1000 simulations of $e_1/\text{sum}(e[k])$ from Wishart$(I, 17, 44)/17$, compared to actual value .207

1000 permutations of 1st eigenvector slope, vs coordinate index, and actual slope .00515
The one-sample $t$ statistic observe

$x_1, x_2, \ldots, x_n \sim \text{N}(\mu, \sigma^2)$ and wish to

Test $H_0: \mu = 0$. For a two-sided test use

\[
T^2 = \frac{||\bar{x}||^2}{||x||^2/(n-1)} = \frac{\bar{x}^2}{(x-x^*^T/\sigma^2)}
\]

\[
= \text{Cov}(\hat{A} \cdot (x-1))
\]

Reject $H_0$ for $x$ closely aligned in either direction, with $1/\infty$.

One-sided test

Reject only for small values of $\hat{A}$ (not of $\hat{A}$ or $\pi - \hat{A}$).

One-sided test

Reject only for small values of $\hat{A}$ (not of $\hat{A}$ or $\pi - \hat{A}$).
Hotelling’s One-Sample $T^2$ Test

Hotelling developed the analogous theory for

\[ x_1, x_2, \ldots, x_n \sim \mathcal{N}_p (\mu, \Sigma) \]

Testing \( H_0: \mu = 0 \cdot \text{rho} \)

The $T^2$ statistic

\[ T^2 = (\bar{x} - \mu_0)' \Sigma^{-1} (\bar{x} - \mu_0) \sim \chi^2_{p} \]

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \sim \mathcal{N}_p (\mu, \Sigma/n) \quad \text{independent of} \]

\[ S = \sum_{i=1}^{n} (x_i - \bar{x}) (x_i - \bar{x})' = \sum_{i=1}^{n} x_i x_i' - n \bar{x} \bar{x}' \sim W(p; \mu, \Sigma, p) \]

Hotelling’s one-sample $T^2$ statistic is

\[ \frac{T^2}{\bar{x}' (\sum_{i=1}^{n} \frac{S}{n-1})^{-1} \bar{x}} \]

Unbiased estimate of \( \frac{\bar{x}'}{\bar{x}'} \circ \text{Cov} \bar{x} \)
Geometric Interpretation Lab

\( \hat{A} \) be the minimum angle between \( \mathbb{L}_2^n \) and any vector in \( \mathbb{L}_v^n (x) \):

\[
\hat{A} = \cos^{-1}(\hat{A}) : (n-1)
\]

Proof. According to property (vii) A2.3

\[
\cos^{-1}(\hat{A}) = \frac{\hat{X}'(XX')^{-1}X}{\sqrt{\hat{X}'X}} = \frac{x'}{\sqrt{x'x}}
\]

= \( x' (S + xx')^{-1} x \).

However by Woodbury's theorem, A4.17,

\[
(S + xx')^{-1} = S' [S - \frac{xx'}{\sqrt{x'x}^2 + x'x}] S'
\]
\[ \cos \hat{A} = n \bar{x}' \tilde{S}' \bar{x} - \frac{(n \bar{x}' \tilde{S}' \bar{x})^2}{n \bar{x}' \tilde{S}' \bar{x}^2} = n \bar{x}' \tilde{S}' \bar{x} \]

Therefore:

\[ \cot^2 \hat{A} = \frac{\cos^2 \hat{A}}{1 - \cos^2 \hat{A}} = n \bar{x}' \tilde{S}' \bar{x} \]

\[
\left( \text{So } (n-1) \cot^2 = n(n-1) \bar{x}' \tilde{S}' \bar{x} = \bar{x}' \frac{\tilde{S}^2 \bar{x}}{\lambda_{\text{row}}} \bar{x} = 1 \right)
\]

Homework 1

\[ \sqrt{v_w} = \frac{u \cdot v}{||u||}, \quad ||v|| = 1, \quad \text{and } A_{u,v} \text{ min angle between } u \text{ and } v \]

Show that:

\[ \cot^2 A_{u,v} = S_{u,v} \frac{1 - \bar{x}'}{S_{v,v} S_{v,u}} \]

when:

\[ S_{v,v} = v \cdot v, \quad S_{u,v} = \bar{x}', \quad S_{v,u} = v (I - uu') v \]

Homework 2

Apply above to case \( p=1 \), and show that you get the one-sample \( t \) test.
Null Distribution $\chi^2$ If $H_0$ is true

is true that $X \sim N(0, \Sigma \otimes I_p)$, and

according to B1.15, $L = L_{row}(X)$ will be uniformly distributed over the set of $p$-dimensional subspaces of $\mathbb{R}^n$.

To find the distribution of $\hat{A}$, it is equivalent to think of $L$ as fixed, while $L$ is replaced by $y = (y_1, \ldots, y_p)$ where

$y \sim N(\eta, I_p)$. Because $y$ will be uniformly distributed in

and $L = L(e_n, e_2, \ldots, e_p)$, $e_1 = (33 - \bar{x}) \ldots (\bar{x})$, so

$$
\text{cov} \hat{A} = \frac{\hat{y}_1^2 + \hat{y}_2^2 + \ldots + \hat{y}_p^2}{\hat{y}_1 + \ldots + \hat{y}_p} \sim \frac{p}{\hat{p} + \hat{\beta}} F_{p, n-p}.\n$$
Therefore the null distribution of the one-sample $T^2$ statistic is

$$T^2 \sim \frac{n-1}{n-p} p F_{n-p}.$$ 

\underline{Data-Sampling} \hspace{1cm} \underline{X = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix}}.

The rows $\underline{y} = (y_1, y_2, \ldots, y_n)$ represent possible response variables $i$ that we might wish to check for significance. For example, each of the $n$ subjects in a drug trial might have his or her blood levels of some drug measured at hours $1, 2, \ldots, p$ after injection.

We might select some particular response $a$ to check, say
have 3 in response, so scientific grounds, and do the one-sample

A test on the n numbers \( x_1, x_2, \ldots, x_n \) in terms of the
picture, this amounts to seeing if the angle \( \hat{A} \) is too small
to have occurred by chance (or \( \min (\hat{A}, \pi - \hat{A}) \) for a
two-sided test.) Or, more aggressively, we might use

\[
\min \left( \frac{\hat{A}}{\pi} , \frac{\pi - \hat{A}}{\pi} \right)
\]

as test statistic, though this makes it hard to find the
null hypothesis distribution.

Fisher's test amounts to using the apparently best
Linear combination of variables, \( \hat{\nu} = \sum_{i} \alpha_{i} \hat{\nu}_{i} \).

**Hawk 3.** Show that Hotelling's one-sample test is equivalent to doing an ordinary two-sided one-sample test with response variable \( \hat{\nu} = \sum_{i} \alpha_{i} \hat{\nu}_{i} \), the projection of \( \hat{\nu} \) into \( \hat{\mathbf{L}}_{i} \).

- **Advantage.** The \( T^{2} \) test automatically "scales" through the data in order to choose the 'best' response variable.

- **Disadvantage.** Screening can be expensive in terms of statistical power.

**Example.** \( X \sim N_{5 \times 50} (0, I_{5} \otimes I_{50}) \). Observe \( \hat{\nu} = X + \mu \)

\( (\mu = 0.01, 0.2, 0.25) \)
* Run 3 tests of $H_0: \mu = 0$.

1. Hotelling's $T^2$:
   
   (2) One-sample two-sided $t$ test for response
   
   \[ T^2 = \frac{1}{s^2} \sum (X_i - \bar{X})^2 \]

2. One-sample one-sided

I ran 200 simulations of the 3 tests, for each value of $\mu = 0, .01, .02, ... .25$, and recorded the achieved significance levels (i.e., $p$-values.) Figure B5 shows the price $T^2$ pays in terms of reduced ASEL.

Tukey's rule. Usually, better to make educated guess for $v$. 
figB5. Simulation comparing Hotelling's one-sample T2 with one and two-sided Mean tests, p=5, n=50
Relation With Univariate \( t \)-test

\[
X \sim \mathcal{N}(\mu, \Sigma, \mathbf{1})
\]

\[
\mathbf{x} \sim \mathcal{N}(\mu, \Sigma, \mathbf{1})
\]

\[
Y = \mathbf{a}^\top \mathbf{x} = (y_1 = \mathbf{a}_1^\top \mathbf{x}, y_2 = \mathbf{a}_2^\top \mathbf{x}, \ldots) \sim \mathcal{N}(\mu = \mathbf{a}, \Sigma = \mathbf{a}^\top \mathbf{a})
\]

One-sample \( t \)-statistic

\[
\hat{t}(\mathbf{a}) = \frac{\mathbf{y}^\top \mathbf{y}}{\mathbf{y}^\top \Sigma^{-1} \mathbf{y}} = \frac{n(n-1)(\mathbf{a}^\top \mathbf{\bar{x}})^2}{\mathbf{a}^\top \Sigma \mathbf{a}}
\]

*Remark* 3: Show that \( \max_{\mathbf{a}} \hat{t}(\mathbf{a})^2 = \frac{1}{2} \)

achieved at \( \mathbf{a} = \hat{\mathbf{a}} \mathbf{X} \mathbf{0} \), where \( \mathbf{0} \) is the projection of \( \mathbf{1} \) into \( \mathbf{L}(\mathbf{X}) \).

So \( t \)-test rejects for \( \hat{t}(\mathbf{a}) > \left[ \frac{(n-1)p}{n-p} \frac{F_{p,n-p}}{\mathbf{1}} \right]^{\frac{1}{2}} \max \text{ possible } \hat{t}\text{-stat} \)

Example: \( n = 21, \alpha = 0.05 \). Here are the rejection values:

<table>
<thead>
<tr>
<th>((n-1)p / \mathbf{1})</th>
<th>2.09</th>
<th>2.72</th>
<th>3.25</th>
<th>3.74</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbf{1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hotelling's Two-Sample Test

- Univariate 2-sample t-test

\[
\begin{cases}
    \bar{x} - \bar{y} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}) \\
    \bar{x} \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{n_1}) \\
    \bar{y} \sim \mathcal{N}(\mu_2, \frac{\sigma^2}{n_2})
\end{cases}
\]

Hypothesis: \( H_0: \mu_1 = \mu_2 \)

- Test statistic

\[ t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

- Distribution

\[ t \sim t_{n_1 + n_2 - 2} \]

and

\[ t = \left( \text{Cot} (\hat{A}) \right)^2 \cdot (\eta - 2) \]

as shown above.
Homework 4/
Verify the boxed statement.

H_0: \mu_1 = \mu_2

\begin{align*}
H_0' & : \rho_1 = \rho_2 \\
\text{Sufficient statistics:} & \quad \bar{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}) (X_i - \bar{X})' \\
S_1 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}) (X_i - \bar{X})' \\
S_2 &= \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y}) (Y_i - \bar{Y})'
\end{align*}

\begin{align*}
\mathbf{T}^2 &= (\bar{Y} - \bar{X})' \left( \frac{1}{n} S_1 + \frac{1}{n-1} S_2 \right)^{-1} (\bar{Y} - \bar{X}) \\
\text{Geometric Interpretation:} & \quad \mathbf{Z} = (X, Y) = \begin{pmatrix} -3 \\ -2 \end{pmatrix} \\
\mathbf{P} &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\
\mathbf{T}^2 &= \cot^2(\hat{A}) \cdot (n-2) \quad \text{where } \hat{A} \text{ is the minimum angle between } \mathbf{P} \text{ and } \mathbf{Z} \end{align*}
Homework 5: Argue from our previous result that the null hypothesis distribution is

\[
\frac{2}{T} \sim \frac{n-2}{n-p-1} \quad F_p, n-p-1
\]

(s. \( H_0 \) is rejected at level \( \alpha \) if \( T > F_{p, n-p-1}^{(\alpha)} \)).

Homework 6: "prostate." is a 6033 x 102 microarray expression matrix; first 50 columns are normal controls, last 52 columns prostate cancer patients. (6) Compute the 2-sample \( T \) statistic

For \( T = \text{prostate\,data}[1:10,\,] \), the first 10 rows (genes). (6) Repeat, now removing row 2. (5) Compute individual two-sample \( t \)-tests for the first 10 genes and say why (6) is so much different than (5).
Non-Null $t^2$ Distribution (For convenience go back to one-sample case)

Invariance

Start with $\tilde{X} \sim \mathcal{N}_{p \times n}(\mu \tilde{I}, \Sigma \otimes I_n)$, and for $A$ non-singular, let

$$\tilde{X} \sim A \tilde{X}$$

Then

$$\tilde{X} = A \tilde{X} \quad \text{and} \quad \tilde{S} = A \tilde{S} A' = ASA'$$

So

$$\tilde{T}^2 = \tilde{X}' A \left[ \frac{ASA'}{\lambda(n-1)} \right] A \tilde{X} = T^2$$

So

$$\tilde{T}^2 \sim \tilde{X} \text{ Normal } \frac{\tilde{X}' A \tilde{X}}{\lambda(n-1)}$$

By taking $A = \hat{\Sigma}^{-\frac{1}{2}}$, so $A \hat{\Sigma} A' = I_p$, we get

$$\tilde{X} \sim \mathcal{N}_{p \times n}(\mu \tilde{I}, I_p \otimes I_n)$$

So we can always work with $\tilde{X}$ to simplify calculations.
Hotelling's Theorem. If \( X \sim N_p(\mu, \Sigma), \Sigma \geq 0 \),

then \( T^2 = \bar{x}' \left( \frac{1}{n-1} \right) \bar{x} \) has distribution

\[ T^2 \sim \frac{n-1}{p} F_{p, n-p}(0, \mu, \Sigma) \]

where \( F_{p, n-p}(0, \mu, \Sigma) \) is the "non-central F distribution".

\[ F_{p, n-p}(0, \mu, \Sigma) \sim \frac{X^2(\delta)}{\mu} / \frac{X^2(n-p)}{n-p} \]

Lemma. If fixed unit vector, \( S \sim W(I, m, p) \) for \( m \geq p \).

Then

\[ WSU \sim \frac{1}{m-p+1} \]

Proof. For and \( \Delta \),
\( L \leq U = U' \Delta S' \Delta U = (\Delta U' \Delta S \Delta')' (\Delta U') \)

\[ \leq \tilde{W}' \tilde{S} \tilde{U} \]

But \( S \sim W(\Delta A'; m, p), \ B3.3(\chi), \) and we can choose \( A \) so that \( \tilde{U} = \Delta U = (0, c, -c, 1)' \), implying \( \tilde{e}_p \)

This gives \( U' \leq U \sim (S')_{pp}, \ S \sim W(I_3; m, p). \)

\[ \frac{U}{V} \sim \frac{S}{W} \]

**Proof of Theorem**: Suppose \( X \sim \mathcal{N}(\mu, \Sigma), \) and \( \tilde{\mu} = \frac{\mu_p}{\mu_1} \)

\[ U = \bar{X}/||\bar{X}|| \]

So

\[ T^2 = ||\bar{X}||^2 \chi^2(p-1), \ U' \leq \tilde{U} \]

**Hand**: Finish the Lemma's Proof. Hint: Consider \( X = (X_1, ..., X_p)' \)

**Proof**
We know that (B.1.6 (iii))

\[
\bar{x} \sim \mathcal{N}_p (\bar{\mu}, \frac{I_p}{n}) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{N}_p (\mu_i, I_p)
\]

independent of \( p \)

\[\sum_{i=1}^{n} \sim \mathcal{W}_p (I_p, n-1, p)\]

 Conditioning on \( \bar{x} \) and using the Lemma with \( m = n-1 \),

\[
T^2 | \bar{x} \sim \frac{(\bar{x} - \bar{\mu}^2)}{\frac{1}{n} \sum_{i=1}^{n} (\mu_i - \bar{\mu})^2} \sim \chi^2_{n-1}
\]

independent of \( \bar{x} \)

\[
\frac{1}{n} \sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 \sim \frac{\chi^2_{n-1}}{\chi^2_{n-1}} \chi^2_{n-1}
\]

\[
\sim \chi^2_{n-1} \sim \frac{\chi^2_{n-1}}{\chi^2_{n-1}} F_{n-1, p, \infty} \sim F_{n-1, p, \infty}
\]
Slightly more general statement \( x \sim \mathscr{N}(\mu, \Sigma) \) if \( S \sim W(p, \Sigma, n) \)

\[
\bar{x}' \left( \frac{S}{n} \right) \sim \frac{mp}{m-p+1} F_{p, m-p+1}(\delta)
\]

\[
S = \frac{\bar{x}' \bar{x}}{\left( \frac{p}{n} \right)}
\]

Problem 8: Show that Hotelling's two-sample \( T^2 \) test is distributed as

\[
\frac{X^2}{\sum_i \frac{(n-2)F_{p, n-p+1}(\delta) + \delta}{p}} \sim \frac{S^2}{\left( \frac{p}{n} \right)}
\]

Mahalobis Distance

The power of the \( T^2 \) test depends on

\[
\Lambda = (\mu_x - \mu_y)' \Sigma^{-1} (\mu_x - \mu_y)
\]
In general if $X$ and $Y$ have mean and covariance

$$X \sim \mathcal{N}_p(\mu_X, \Sigma_X) \text{ and } Y \sim \mathcal{N}_p(\mu_Y, \Sigma_Y),$$

not necessarily normal, then

$$\Delta = \left( \frac{\mu_X - \mu_Y}{\sqrt{\Sigma_X}} \right)^T \Sigma_Y^{-1} \left( \frac{\mu_X - \mu_Y}{\sqrt{\Sigma_X}} \right)^{\frac{1}{2}}$$

is called the Mahalanobis distance between the distributions of $X$ and $Y$. [This definition does not require $X$ and $Y$

to be independent, or even to be defined on the same space, i.e., $\Delta = \left( \frac{\mu_X}{\sqrt{\Sigma_X}} \right)^T \Sigma_Y^{-1} \Delta (x)$ is the Mahalanobis distance from $\mathcal{N}_p(\mu_Y, \Sigma_Y)$ to $\mathcal{N}_p(\mu_X, \Sigma_X)$.

2. 

Hint: If $X \sim \mathcal{N}_p(\mu_X, \Sigma_X)$ and $Y \sim \mathcal{N}_p(\mu_Y, \Sigma_Y)$, then show that

the Kullback-Leibner directed divergence

$$\int_{\mathbb{R}^p} f_X(x) \log \frac{f_X(x)}{f_Y(x)} \, dx = 2 \Gamma^2,$$
Partitions \( \Delta^2 \)

\[
X = \left( \begin{array}{c} X_1 \\ X_2 \end{array} \right) = \left( \begin{array}{c} \mu' \\ \mu_2 \end{array} \right), \quad \left( \begin{array}{c} \frac{x_1}{2} \\ \frac{x_2}{2} \end{array} \right)
\]

Define

\[
\mu_2 = \frac{1}{2} - \frac{X_1}{2} \frac{X_2}{2} \quad \text{and} \quad \frac{X_2}{2} = \frac{X_2}{2} - \frac{X_2}{2} \frac{X_2}{2}.
\]

Then

\[
\Delta^2 (X) = \mu' \frac{X_2}{2} \mu = \mu_1 \frac{X_2}{2} \mu_2 + \frac{1}{2} \left( \frac{X_2}{2} \right)^2 / 2.
\]

**Homework 10**

(a) Verify \( \Delta \)

\( \text{Hint: First consider} \)

\[
X_1 \sim N_1 (\mu_1, \Sigma_1) \quad \text{and} \quad X_2 = X_2 - \frac{X_1}{2} \frac{X_2}{2} \sim N_2 (\mu_2, \Sigma_2).
\]

Then use the "preservation property".

(b) How would you test \( H_0 : \Delta = \Delta_0 = \mu' \frac{X_2}{2} \mu \)?

**Preservation Property** \( X = AX \), \( A \) non-singular: \( \hat{\mu} = A \mu, \hat{X} = A \Sigma A^T \)

and

\[
\hat{\Delta} = \mu' A^T (A \Sigma A^T) A \mu = \Delta.
\]