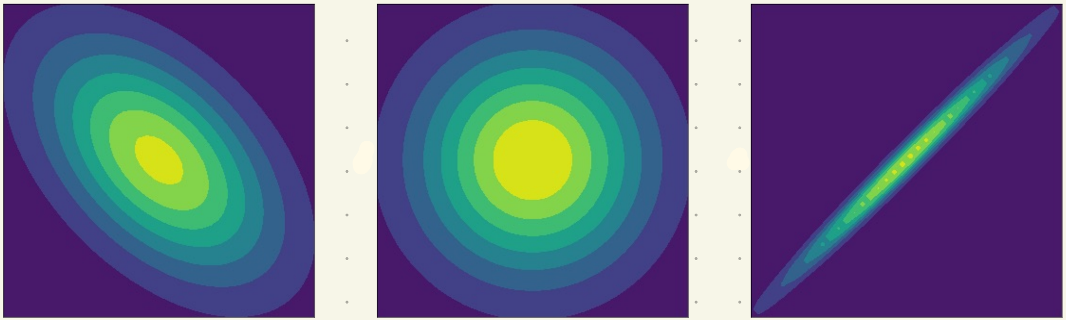
Review:
MVN distribution



https://towardsdatascience.com/no-stress-gaussian-processes-40e238597864

Multivariate normal (MVN) distribution

A random vector $x \in \mathbb{R}^p$ has p-dimensional MVN distribution iif $v^T x$ is a univariate normal random variable for all $v \in \mathbb{R}^p$.

Denote $x \sim N_p(\mu, \Sigma)$, where $\underset{\text{mean}}{\mu \in \mathbb{R}^p}$ and $\underset{\text{covariance}}{\Sigma \in \mathbb{R}^{p \times p}}$

Properties

- If x is p-variate normal and $A \in \mathbb{R}^{q \times p}$, $b \in \mathbb{R}^q$ then $Ax + b$ is q-variate normal

  | $v^T(Ax+b) = \tilde{v}^T x + \tilde{b}$

- If $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ is p-variate normal then x_i are univariate normal

  | $e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ ith $\quad e_i^T x = x_i$

MVN density

If $\Sigma > 0$ then the density for $x \sim N_p(\mu, \Sigma)$ is

$$f(x) = \left(\frac{1}{2\pi}\right)^{p/2} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

and the characteristic function is

$$\varphi_x(t) = e^{it^T\mu - \frac{1}{2}t^T \Sigma t}$$

Transformation

- If $x \sim N_p(\mu, \Sigma)$ then $y = Ax + b$ where $A \in \mathbb{R}^{q \times p}$, $b \in \mathbb{R}^q$ has distribution $y \sim N_q(A\mu + b,\ A\Sigma A^T)$.

  $E(y) = A\mu + b$, $\mathrm{cov}(y) = A\Sigma A^T$

- If $x \sim N_p(0, I_p)$ and $y = \Sigma^{1/2}x + \mu$ then $y \sim N_p(\mu, \Sigma)$

- If $x \sim N_p(\mu, \Sigma)$ where Σ is full rank then
$$y = \Sigma^{-1/2}(x - \mu) \sim N_p(0, I_p)$$

Independence

- If $\begin{pmatrix} x \\ y \end{pmatrix}$ is MVN then

  x and y are independent $\iff$ x and y are uncorrelated

  This is <u>not</u> true for a general distribution.

- If $\begin{pmatrix} x \\ y \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{pmatrix}\right)$ then

  x and y are independent and

  $x \sim N_p(\mu_x, \Sigma_x)$ and $y \sim N_q(\mu_y, \Sigma_y)$

  | density functions and $|\Sigma| = |\Sigma_x| \cdot |\Sigma_y|$

- If $x \sim N_p(\mu, \Sigma)$ then Ax and Bx are independent

  iff $A \Sigma B^T = 0$

  | $\mathrm{cov}(Ax, Bx) = A \,\mathrm{cov}(x)\, B^T = 0$

Conditional distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}\right),$$ denote $z = y - \Sigma_{yx}\Sigma_x^{-1}x$

- x and z are normal

$$x = \underbrace{(I \ \ 0)}_{A}\begin{pmatrix} x \\ y \end{pmatrix} \qquad z = \underbrace{(-\Sigma_{yx}\Sigma_x^{-1}, I)}_{B}\begin{pmatrix} x \\ y \end{pmatrix}$$

- $x \sim N_p(\mu_x, \Sigma_x)$

$$z \sim N_q(\mu_y - \Sigma_{yx}\Sigma_x^{-1}\mu_x, \ \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy})$$

$$E(Bx) = B \cdot \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

$$\mathrm{cov}(Bx) = B\Sigma B^T = (-\Sigma_{yx}\Sigma_x^{-1}, I)\begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}\begin{pmatrix} -\Sigma_x^{-1}\Sigma_{xy} \\ I \end{pmatrix}$$

$$= (\underbrace{-\Sigma_{yx} + \Sigma_{yx}}_{=0}, \ -\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy} + \Sigma_y)\begin{pmatrix} -\Sigma_x^{-1}\Sigma_{xy} \\ I \end{pmatrix}$$

- x and z are independent

$$A \Sigma B^T = (I\ \ 0)\begin{pmatrix}\Sigma_x & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_y\end{pmatrix}\begin{pmatrix}-\Sigma_x^{-1}\Sigma_{xy}\\ I\end{pmatrix} = (\Sigma_x\ \ \Sigma_{xy})\begin{pmatrix}-\Sigma_x^{-1}\Sigma_{xy}\\ I\end{pmatrix} =$$
$$= -\Sigma_{xy} + \Sigma_{xy}$$

- $y|x \sim N_q\left(\mu_y + \Sigma_{yx}\Sigma_x^{-1}(x-\mu_x),\ \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\right)$

$$z = y - \Sigma_{yx}\Sigma_x^{-1}x \Rightarrow y = z + \Sigma_{yx}\Sigma_x^{-1}x$$

independent

$$E(y|x) = E(z) + \Sigma_{yx}\Sigma_x^{-1}x = \mu_y - \Sigma_{yx}\Sigma_x^{-1}\mu_x + \Sigma_{yx}\Sigma_x^{-1}x$$
$$Cov(y|x) = cov(z)$$

Connection to regression

- $z = y - \Sigma_{yx}^{-1}\Sigma_x \cdot X$ represents residuals

$$e = y - \hat{y} = y - XS_x^{-1}S_{xy} \qquad e_i = y_i - S_{yx}S_x^{-1}x_i$$

- $cov(z)$ represents MSE

$$\frac{RSS}{n} = \frac{1}{n}\|e\|^2 = \frac{1}{n}\|y - XS_x^{-1}S_{xy}\|^2 = S_y - S_{yx}S_x^{-1}S_{xy}$$

Geometry of MVN

Given $x \sim N_p(\mu, \Sigma)$, $f(x)$ is the same for all $x \in \mathbb{R}^p$ such that $\underbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}_{}= C$.

$d(x,\mu) =$ Mahalonobis distance from x to μ.

Density of MVN is constant on contours of an ellipsoid.

$\Sigma = U \Lambda U^T$, denote $y = U^T(x-\mu) \sim N_p(0, \Lambda)$

Contours for y: $y^T \Lambda^{-1} y = c \iff \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = c$ ellipsoid

$x = Uy + \mu$ (rotation + shift)

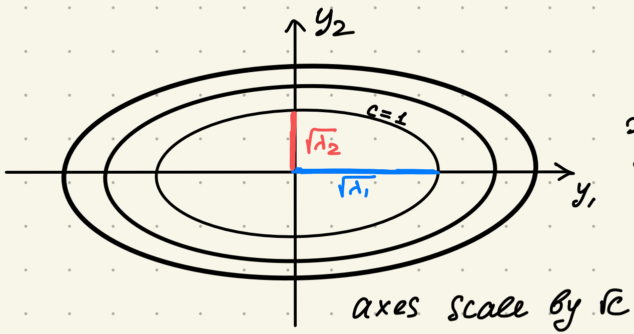
$\boxed{p=2}$ $$x \sim N_2(\mu, \Sigma)$$

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

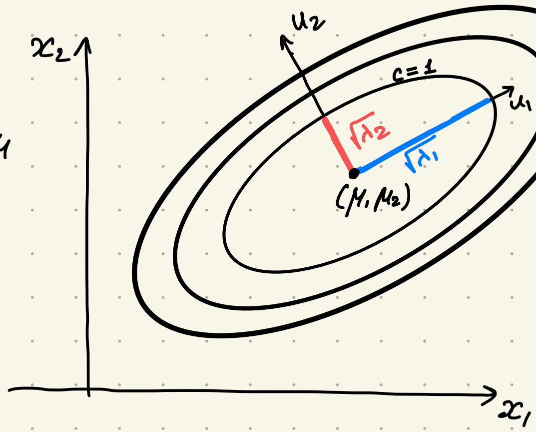
$$\Sigma = U \Lambda U^T \qquad y = U^T(x - \mu) \sim N(0, \Lambda)$$

$$\begin{pmatrix} | & | \\ u_1 & u_2 \\ | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} - u_1^T - \\ - u_2^T - \end{pmatrix}$$

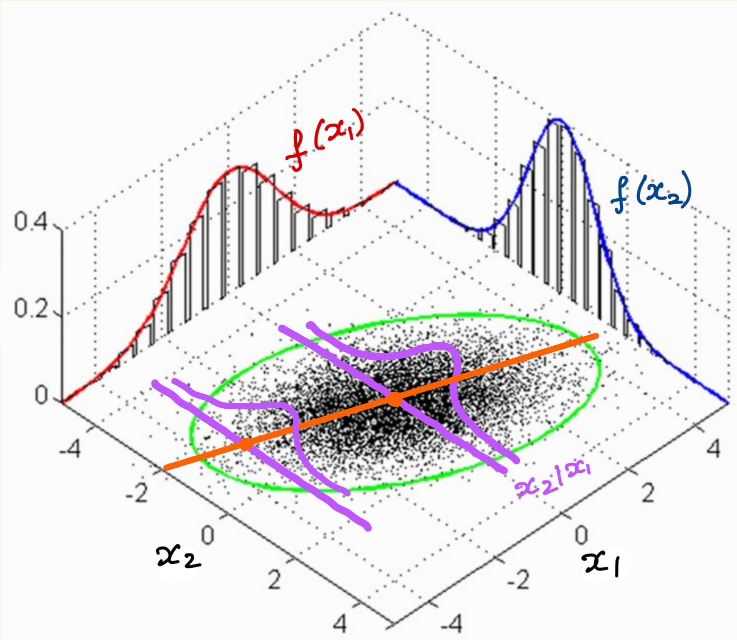
Contours for y are $\dfrac{y_1^2}{\lambda_1} + \dfrac{y_2^2}{\lambda_2} = c$



$x = Uy + \mu$



$y|x \sim N_q(\mu_y + \Sigma_{yx}\Sigma_x^{-1}(x-\mu_x),\ \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy})$



orange line: $\mu_2 + \frac{\Sigma_{21}}{\Sigma_{11}}(x_1 - \mu_1)$

- it passes through (μ_1, μ_2), how about the direction?
- variance of $x_2|x_1$ is $\Sigma_{22} - \frac{\Sigma_{12}\Sigma_{21}}{\Sigma_{11}}$

Confidence ellipses

If $x \sim N_p(\mu, \Sigma)$ and $\Sigma > 0$ then

$$(x-\mu)^T \Sigma^{-1} (x-\mu) \sim \chi^2(p)$$

This allows to compute $\mathbb{P}[(x-\mu)^T \Sigma^{-1}(x-\mu) < t]$

and draw confidence ellipses (e.g. 95%)

$$y = \Sigma^{-1/2}(x-\mu) \sim N_p(0, I)$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \quad \text{and} \quad y_i \overset{iid}{\sim} N(0,1)$$

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = y^T y = \sum_{i=1}^{p} y_i^2 \sim \chi^2(p)$$

Schur complement

$M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \mathbb{R}^{(p+q)\times(p\times q)}$ $\quad A \in \mathbb{R}^{p\times p}$, $C \in \mathbb{R}^{q\times q}$ are invertible

Schur complement of C is $M/C = A - BC^{-1}B^T \in \mathbb{R}^{p\times p}$

of A is $M/A = C - B^TA^{-1}B \in \mathbb{R}^{q\times q}$

Properties

- $M \succ 0$ iif $A \succ 0$ and $M/A \succ 0$
- $M \succ 0$ iif $C \succ 0$ and $M/C \succ 0$

Block inverse

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}^T & \tilde{C} \end{pmatrix} = \begin{pmatrix} (M/C)^{-1} & -A^{-1}B\,(M/A)^{-1} \\ -C^{-1}B^T(M/C)^{-1} & (M/A)^{-1} \end{pmatrix}$$

$$\begin{cases} A\tilde{A} + B\cdot\tilde{B}^T = I \\ A\tilde{B} + B\cdot\tilde{C} = 0 \Rightarrow \tilde{B} = -A^{-1}B\tilde{C} \\ B^T\tilde{A} + C\tilde{B}^T = 0 \\ B^T\tilde{B} + C\tilde{C} = I \Rightarrow -B^TA^{-1}B\tilde{C} + C\tilde{C} = I \end{cases}$$

$$\Rightarrow \tilde{C} = (C - B^TA^{-1}B)^{-1}$$

By analogy

$$\tilde{A} = (A - BC^{-1}B^T)^{-1}$$

$$\tilde{B}^T = -C^{-1}B^T\tilde{A}$$

Partial correlation

$$x = \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix} \sim N_{p_1+p_2}\left(\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \qquad (p = p_1 + p_2)$$

$$x_{(2)} \mid x_{(1)} \sim N_{p_2}\left(\mu_{(2)} + \Sigma_{21} \Sigma_{11}^{-1} (x_{(1)} - \mu_{(1)}),\ \underbrace{\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}}_{\Sigma_{2\cdot 1}} \right)$$

$\Sigma_{2\cdot 1}$ is called partial covariance.

- $\Sigma_{2\cdot 1}$ is Schur complement Σ / Σ_{11}

Now, let $x_{(2)} = \begin{pmatrix} x_i \\ x_j \end{pmatrix}$ then $\Sigma_{2\cdot 1} \in \mathbb{R}^{2\times 2}$ and could be used to compute partial correlation

$$\rho_{ij \mid 12 \ldots (i-1)(i+1) \ldots (j-1)(j+1) \ldots p}$$

Consider the precision matrix

$$\Theta = \Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} * & * \\ * & \underbrace{(\Sigma / \Sigma_1)^{-1}}_{\text{inverse of partial covariance}} \end{pmatrix}$$

If $\Sigma / \Sigma_{11} = \Sigma_{2\cdot 1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ $(a_{21} = a_{12})$

then $(\Sigma / \Sigma_{11})^{-1} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \dfrac{1}{a_{11} a_{22} - a_{12} \cdot a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$

Thus $-\dfrac{b_{12}}{\sqrt{b_{11} b_{22}}} = \dfrac{a_{12}}{\sqrt{a_{11} a_{22}}} = \rho_{ij \mid 12 \ldots (i-1)(i+1) \ldots (j-1)(j+1) \ldots p}$

$$\Theta = \Sigma^{-1} = \begin{array}{cc} \begin{pmatrix} \quad & \quad \\ & \boxed{} \end{pmatrix} & \begin{matrix} i \\ j \end{matrix} \\ \quad\quad\; i \;\; j & \end{array}$$

deliver partial correlations

If $\Theta_{ij} = 0$ then x_i and x_j are independent given the rest.

$$(\Sigma/\Sigma_{11})^{-1} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \frac{1}{a_{11}a_{22} - a_{12}\cdot a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

implies $$b_{11} = \frac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}}$$

$$b_{22} = \frac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}}$$

$$b_{12} = -\frac{a_{12}}{a_{11}a_{22} - a_{12}a_{21}}$$

Thus $$-\frac{b_{12}}{\sqrt{b_{11}b_{22}}} = \frac{a_{12}/\cancel{(a_{11}a_{22}-a_{12}a_{21})}}{\sqrt{a_{22}/\cancel{(a_{11}a_{22}-a_{12}a_{21})} \cdot a_{11}/\cancel{(a_{11}a_{22}-a_{12}a_{21})}}} = \frac{a_{12}}{\sqrt{a_{22}a_{11}}}$$

$\Sigma_{2.1}$ can be viewed as covariance matrix and

$$\rho_{ij|12\ldots(i-1)(i+1)\ldots(j-1)(j+1)\ldots p} = \frac{a_{12}}{\sqrt{a_{11}a_{22}}} = -\frac{b_{12}}{\sqrt{b_{11}b_{22}}}$$

This quantity can be computed from θ.

Log-likelihood

If $x_1 \ldots x_n \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)$ then log-likelihood is

$$\ell(X; \mu, \Sigma) = -\frac{n \cdot p}{2} \log(2\pi) - \frac{n}{2} \log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

Now $\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) =$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^T \Sigma^{-1}(x-\bar{x}) + n(\bar{x}-\mu)^T \Sigma^{-1}(\bar{x}-\mu)$$

$$\ell(X; \mu, \Sigma) = \left(\frac{n}{2}(\bar{x}-\mu)^T \Sigma^{-1}(\bar{x}-\mu)\right) + \ldots\ldots$$

$$\nabla_\mu \ell(X; \mu, \Sigma) = n \cdot \underset{0}{\underset{\curlyvee}{\Sigma}}^{-1}(\bar{x}-\mu) = 0 \implies \boxed{\hat{\mu} = \bar{x}}$$

$$\ell(x;\hat{\mu},\Sigma) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x}) + \dots$$

$$= -\frac{n}{2}\log|\Sigma| - \frac{n}{2}\operatorname{tr}(\Sigma^{-1}S) + \dots$$

where $S = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T$

$$\sum_{i=1}^{n}(x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x}) = \sum_{i=1}^{n}\operatorname{tr}[(x_i-\bar{x})\Sigma^{-1}(x_i-\bar{x})^T] =$$

$$= \operatorname{tr}[\Sigma^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})] = n\operatorname{tr}(\Sigma^{-1}S)$$

One can show that $\boxed{\hat{\Sigma} = S}$.

$$\theta = \Sigma^{-1} \Rightarrow \ell(x;\hat{\mu},\theta) = \frac{n}{2}\log|\theta| - \frac{n}{2}\operatorname{tr}(\theta S)$$

$$\nabla_\theta \ell(x;\hat{\mu},\theta) = \frac{n}{2}(\theta^{-1} - S) \Rightarrow \hat{\theta} = S^{-1}$$

Here we used the fact that $\nabla_A \operatorname{logdet}(A) = A^{-T}$

Distribution of the estimates

If $x_1 \dots x_n \overset{iid}{\sim} N_p(\mu, \Sigma)$ then $\bar{x} \sim N_p(\mu, \frac{\Sigma}{n})$

| $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = Ay$ where y is MVN.

If $x_1 \dots x_n \sim N(0, \Sigma)$ then $M = X^T X = \sum_{i=1}^{n} x_i x_i^T \in \mathbb{R}^{p \times p}$ has a Wishart distribution with n degrees-of-freedom and scale matrix Σ. Denote by $M \sim W_p(\Sigma, n)$.

If $\Sigma = I$ the it is standard Wishart distribution

One can show that

$$(n-1)S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \sim W_p(\Sigma, n-1)$$

and independent of $\bar{x}$.

Properties

- $W_p(\Sigma, n)$ is a probability distribution on PSD matrices of size $p \times p$.

- It is a generalization of χ_n^2, if $p=1$ $W_1(\sigma^2, n) \equiv \sigma^2 \chi_n^2$

  | $p=1$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p$, $X^T X = \sum_{i=1}^{p} x_i^2$

- If $M \sim W_p(\Sigma, n)$ and $A \in \mathbb{R}^{p \times q}$ then

  $A^T M A \sim W_p(A^T \Sigma A, n)$

  | $A^T M A = (XA)^T (XA)$, $y = XA = \begin{pmatrix} - y_1^T - \\ \vdots \\ - y_n^T - \end{pmatrix}$, $y_i \sim N_q(0, A^T \Sigma A)$

- If $M \sim W_p(I, n)$ then $\mathrm{tr}(M) \sim \chi^2(np)$

  | $\mathrm{tr} M = \mathrm{tr}\left(\sum_{i=1}^{n} x_i x_i^T\right) = \sum_{i=1}^{n} \mathrm{tr}(x_i^T x_i) = \sum_{i=1}^{n} \|x_i\|^2 \sim \sum_{i=1}^{n} \chi^2(p)$

If $x \sim N_p(0, I)$ and $M \sim W_p(I, n)$ are indenendent $\tau^2 = n\, x^T M^{-1} x$ has Hotelling T^2 distribution with parameters p and n. Denote $\tau^2 \sim T^2(p, n)$.

Properties

- It is a generalization of Student t-distribution
- If $x_1 \dots x_n \overset{iid}{\sim} N_p(\mu, \Sigma)$, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $S = \frac{1}{n-1} \sum_{i=1}^{n} x_i x_i^T$

  $n(\bar{x} - \mu)^T S^{-1} (\bar{x} - \mu) \sim T^2(p, n-1)$

$$\left| \begin{array}{l} \sqrt{n}(\bar{x} - \mu) \sim N_p(0, \Sigma) \\ (n-1) S \sim W_p(\Sigma, n-1) \end{array} \right.$$

– independent

Univariate $(p=1)$	Multivariate
$x_1 \dots x_n \sim N(\mu, \sigma^2)$	$x_1 \dots x_n \sim N_p(\mu, \Sigma)$
$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{x} \sim N_p(\mu, \frac{\Sigma}{n})$
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$ $(n-1)s^2 \sim \sigma^2 \chi^2(n-1)$	$S = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$ $(n-1)S \sim W_p(\Sigma, n-1)$
$\frac{(\bar{x}-\mu)^2}{s^2/n} \sim t^2(n-1)$	$n(\bar{x}-\mu)S^{-1}(\bar{x}-\mu) \sim T_p^2(n-1)$

Recap from Lecture 2

x is p-dimensional MVN if $v^T x$ is univariate normal for any $v \in \mathbb{R}^p$.

- density for $x \sim N_p(\mu, \Sigma)$ is

$$f(x) = \left(\frac{1}{2\pi}\right)^{p/2} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2} \underbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}}$$

- density contours are ellipsoids $(x-\mu)^T \Sigma^{-1} (x-\mu) = c$

Properties

- $x \sim N_p(\mu, \Sigma)$ then $\overset{q}{y} = \overset{q \times p}{A} \overset{p}{x} + \overset{q}{b} \sim N_q(A\mu + b, A \Sigma A^T)$,

  e.g. $y = \Sigma^{-1/2}(x-\mu) \sim N_p(0, I)$

- $\binom{x}{y}$ is MVN

  x and y are independent $\Longleftrightarrow$ x and y are uncorrelated

- $\begin{pmatrix} x \\ y \end{pmatrix}^{\in \mathbb{R}^p}_{\in \mathbb{R}^q} \sim N_{p+q}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix} \right)$ then

  $y|x \sim N_q\left(\underbrace{\mu_y + \Sigma_{yx} \Sigma_x^{-1} (x - \mu_x)}_{\text{linear in } x}, \underbrace{\Sigma_y - \Sigma_{yx} \Sigma_x^{-1} \Sigma_{xy}}_{\text{Constant}} \right)$

- If $\Theta_{ij} = 0$ in the precision matrix $\Theta = \Sigma^{-1}$ then x_i and x_j are independent given $x_1 \dots \cancel{x_i} \dots \cancel{x_j} \dots x_n$

- If $x_1 \dots x_n \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)$ then MLE are

  $\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$

- Distribution of the estimates

  $\bar{x} \sim N_p\left(\mu, \frac{\Sigma}{n}\right)$ and $\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \sim W_p(\Sigma, n-1)$