Likelihood Methods of Inference

Given data X with model $\{f_\theta(x); \theta \in \Theta\}$:

Definition: The likelihood function is map L: domain Θ, values given by

$$L(\theta) = f_\theta(X)$$

Key Point: think about how the density depends on θ not about how it depends on X.

Notice: X, observed value of the data, has been plugged into the formula for density.

We use likelihood for most inference problems:

1. Point estimation: we must compute an estimate $\hat{\theta} = \hat{\theta}(X)$ which lies in Θ. The **maximum likelihood estimate (MLE)** of θ is the value $\hat{\theta}$ which maximizes $L(\theta)$ over $\theta \in \Theta$ if such a $\hat{\theta}$ exists.

2. Point estimation of a function of θ: we must compute an estimate $\hat{\phi} = \hat{\phi}(X)$ of $\phi = g(\theta)$. We use $\hat{\phi} = g(\hat{\theta})$ where $\hat{\theta}$ is the MLE of θ.

3. Interval (or set) estimation. We must compute a set $C = C(X)$ in Θ which we think will contain θ_0. We will use

$$\{\theta \in \Theta : L(\theta) > c\}$$

for a suitable c.

4. Hypothesis testing: decide whether or not $\theta_0 \in \Theta_0$ where $\Theta_0 \subset \Theta$. We base our decision on the likelihood ratio

$$\frac{\sup\{L(\theta); \theta \in \Theta_0\}}{\sup\{L(\theta); \theta \in \Theta \setminus \Theta_0\}}$$

Maximum Likelihood Estimation

To find MLE maximize L.

Typical function maximization problem:

Set gradient of L equal to 0

Check root is maximum, not minimum or saddle point.

Often L is product of n terms (given n independent observations).

Much easier to work with logarithm of L: log of product is sum and logarithm is monotone increasing.

Definition: The **Log Likelihood** function is

$$\ell(\theta) = \log\{L(\theta)\}\,.$$

Samples from MVN Population

Simplest problem: collect replicate measurements $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from single population.

Model: X_i are iid $MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Parameters (θ): $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Parameter space: $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}$ is some positive definite $p \times p$ matrix.

Log likelihood is

$$\begin{aligned}\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -np\log(\pi)/2 - n\log\det\boldsymbol{\Sigma}/2 \\ - \sum(\mathbf{X}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})/2\end{aligned}$$

Take derivatives.

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\mu}} &= \boldsymbol{\Sigma}^{-1}\left\{\sum(\mathbf{X}_i - \boldsymbol{\mu})\right\} \\ &= n\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})\end{aligned}$$

where $\bar{\mathbf{X}} = \sum \mathbf{X}_i/n$.

Second derivative wrt $\boldsymbol{\mu}$ is a matrix:

$$-n\boldsymbol{\Sigma}^{-1}$$

Fact: if second derivative matrix is negative definite at critical point then critical point is a maximum.

Fact: if second derivative matrix is negative definite everywhere then function is concave; no more than 1 critical point.

Summary: ℓ is maximized at

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$$

(regardless of choice of $\boldsymbol{\Sigma}$).

More difficult: differentiate ℓ wrt $\boldsymbol{\Sigma}$.

Somewhat simpler: set $\mathbf{D} = \boldsymbol{\Sigma}^{-1}$

First derivative wrt $\mathbf{D}$ is matrix with entries

$$\frac{\partial \ell}{\partial \mathbf{D}_{ij}}$$

Warning: method used ignores symmetry of $\boldsymbol{\Sigma}$.

Need: derivative of two functions:

$$\frac{\partial \log \det \mathbf{A}}{\partial \mathbf{A}} = \mathbf{A}^{-1}$$

and

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{A}} = \mathbf{x}\mathbf{x}^T$$

Fact: ij^{th} entry of $\mathbf{A}^{-1}$ is

$$(-1)^{i+j} \frac{\det(\mathbf{A}^{(ij)})}{\det \mathbf{A}}$$

where $\mathbf{A}^{(ij)}$ denotes matrix obtained from $\mathbf{A}$ by removing column j and row i.

Fact: $\det(\mathbf{A}) = \sum_k (-1)^{i+k} A_{ik} \det(\mathbf{A}^{(ik)})$; expansion by minors.

Conclusion

$$\frac{\partial \log \det \mathbf{A}}{\partial A_{ij}} = (\mathbf{A}^{-1})_{ij}$$

and

$$\frac{\partial \log \det \mathbf{A}^{-1}}{\partial A_{ij}} = -(\mathbf{A}^{-1})_{ij}$$

Implication

$$\frac{\partial \ell}{\partial \mathbf{D}} = -n\mathbf{\Sigma}/2 - \sum_i (\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T/2$$

Set $=$ 0 and find only critical point is

$$\hat{\mathbf{\Sigma}} = \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T/n$$

Usual sample covariance matrix is

$$\mathbf{S} = \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T/(n-1)$$

Properties of MLEs:

1) $\bar{\mathbf{X}} \sim MVN_p(\boldsymbol{\mu}, n^{-1}\mathbf{\Sigma})$

2) E$(\mathbf{S}) = \mathbf{\Sigma}$.

Distribution of $\mathbf{S}$? Joint distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$?

Univariate Normal samples: Distribution Theory

Theorem: Suppose $X_1, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ random variables. Then

1. $\bar{X}$ (sample mean)and s^2 (sample variance) independent.

2. $n^{1/2}(\bar{X} - \mu)/\sigma \sim N(0, 1)$.

3. $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$.

4. $n^{1/2}(\bar{X} - \mu)/s \sim t_{n-1}$.

Proof: Let $Z_i = (X_i - \mu)/\sigma$.

Then $Z_1, \ldots, Z_p$ are independent $N(0,1)$.

So $Z = (Z_1, \ldots, Z_p)^T$ is multivariate standard normal.

Note that $\bar{X} = \sigma\bar{Z} + \mu$ and $s^2 = \sum(X_i - \bar{X})^2/(n-1) = \sigma^2 \sum(Z_i - \bar{Z})^2/(n-1)$ Thus

$$\frac{n^{1/2}(\bar{X} - \mu)}{\sigma} = n^{1/2}\bar{Z}$$

$$\frac{(n-1)s^2}{\sigma^2} = \sum(Z_i - \bar{Z})^2$$

and

$$T = \frac{n^{1/2}(\bar{X} - \mu)}{s} = \frac{n^{1/2}\bar{Z}}{s_Z}$$

where $(n-1)s_Z^2 = \sum(Z_i - \bar{Z})^2$.

So: reduced to $\mu = 0$ and $\sigma = 1$.

Step 1: Define

$$Y = (\sqrt{n}\bar{Z}, Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z})^T .$$

(So Y has dimension $n+1$.) Now

$$Y = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

or letting $\mathbf{M}$ denote the matrix

$$Y = \mathbf{M}Z .$$

It follows that $Y \sim MVN(0, \mathbf{M}\mathbf{M}^T)$ so we need to compute $\mathbf{M}\mathbf{M}^T$:

$$\mathbf{M}\mathbf{M}^T = \left[\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & -\frac{1}{n} & \ddots & \cdots & -\frac{1}{n} \\ 0 & \vdots & \cdots & & 1-\frac{1}{n} \end{array}\right]$$

$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline & \\ 0 & \mathbf{Q} \end{array}\right] .$$

Put $\mathbf{Y}_2 = (Y_2, \ldots, Y_{n+1})$. Since

$$\text{Cov}(Y_1, \mathbf{Y}_2) = 0$$

conclude Y_1 and $\mathbf{Y}_2$ are independent and each is normal.

Thus $\sqrt{n}\bar{Z}$ is independent of $Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z}$.

Since s_Z^2 is a function of $Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z}$ we see that $\sqrt{n}\bar{Z}$ and s_Z^2 are independent.

Also, see $\sqrt{n}\bar{Z} \sim N(0, 1)$.

First 2 parts done.

Consider $(n-1)s^2/\sigma^2 = \mathbf{Y}_2^T\mathbf{Y}_2$. Note that $\mathbf{Y}_2 \sim MVN(0, \mathbf{Q})$.

Now: distribution of quadratic forms:

Suppose $Z \sim MVN(0, \mathbf{I})$ and $\mathbf{A}$ is symmetric. Put $\mathbf{A} = \mathbf{PDP}^T$ for $\mathbf{D}$ diagonal, $\mathbf{P}$ orthogonal.

Then

$$\mathbf{Z}^T\mathbf{A}\mathbf{Z} = (\mathbf{Z}^*)^T\mathbf{D}\mathbf{Z}^*$$

where

$$\mathbf{Z}^* = \mathbf{P}^T\mathbf{Z}$$

But $\mathbf{Z}^* \sim MVN(0, \mathbf{P}^T\mathbf{P} = \mathbf{I})$ is standard multivariate normal.

So: $\mathbf{Z}^T\mathbf{A}\mathbf{Z}$ has same distribution as

$$\sum_i \lambda_i Z_i^2$$

where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $\mathbf{A}$.

Special case: if all λ_i are either 0 or 1 then $\mathbf{Z}^T\mathbf{A}\mathbf{Z}$ has a chi-squared distribution with df = number of λ_i equal to 1.

When are eigenvalues all 1 or 0?

Answer: if and only if $\mathbf{A}$ is idempotent.

1) If $\mathbf{A}$ idempotent and λ, x is an eigenpair the

$$\mathbf{A}x = \lambda x$$

and

$$\mathbf{A}x = \mathbf{A}\mathbf{A}x = \lambda \mathbf{A}x = \lambda^2 x$$

so

$$(\lambda - \lambda^2)x = 0$$

proving λ is 0 or 1.

2) Conversely if all eigenvalues of $\mathbf{A}$ are 0 or 1 then $\mathbf{D}$ has 1s and 0s on diagonal so

$$\mathbf{D}^2 = \mathbf{D}$$

and

$$\mathbf{A}\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T\mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{P}\mathbf{D}^2\mathbf{P}^t = \mathbf{P}\mathbf{D}\mathbf{P} = \mathbf{A}$$

Next case: $\mathbf{X} \sim MVN_p(0, \mathbf{\Sigma})$. Then $\mathbf{X} = \mathbf{AZ}$ with $\mathbf{AA}^T = \mathbf{\Sigma}$.

Since $\mathbf{X}^T\mathbf{X} = \mathbf{Z}^T\mathbf{A}^T\mathbf{AZ}$ it has the law

$$\sum \lambda_i Z_i^2$$

λ_i are eigenvalues of $\mathbf{A}^T\mathbf{A}$. But

$$\mathbf{A}^T\mathbf{A}x = \lambda x$$

implies

$$\mathbf{AA}^T\mathbf{A}x = \mathbf{\Sigma}\mathbf{A}x = \lambda\mathbf{A}x$$

So eigenvalues are those of $\mathbf{\Sigma}$ and $\mathbf{X}^T\mathbf{X}$ is χ^2_ν iff $\mathbf{\Sigma}$ is idempotent and trace$(\mathbf{\Sigma}) = \nu$.

Our case: $\mathbf{A} = \mathbf{Q} = \mathbf{I} - \mathbf{1}\mathbf{1}^T/n$. Check $\mathbf{Q}^2 = \mathbf{Q}$. How many degrees of freedom: trace($\mathbf{D}$).

Defn: The trace of a square matrix $\mathbf{A}$ is

$$\text{trace}(\mathbf{A}) = \sum \mathbf{A}_{ii}$$

Property: trace($\mathbf{AB}$) = trace($\mathbf{BA}$).

So:

$$\begin{aligned}\text{trace}(\mathbf{A}) &= \text{trace}(\mathbf{PDP}^T) \\ &= \text{trace}(\mathbf{DP}^T\mathbf{P}) = \text{trace}(\mathbf{D})\end{aligned}$$

Conclusion: df for $(n-1)s^2/\sigma^2$ is

$$\text{trace}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) = n - 1.$$

Derivation of the χ^2 density:

Suppose $Z_1, \ldots, Z_n$ independent $N(0,1)$. Define χ^2_n distribution to be that of $U = Z_1^2 + \cdots + Z_n^2$. Define angles $\theta_1, \ldots, \theta_{n-1}$ by

$$\begin{aligned}
Z_1 &= U^{1/2} \cos\theta_1 \\
Z_2 &= U^{1/2} \sin\theta_1 \cos\theta_2 \\
\vdots &= \vdots \\
Z_{n-1} &= U^{1/2} \sin\theta_1 \cdots \sin\theta_{n-2} \cos\theta_{n-1} \\
Z_n &= U^{1/2} \sin\theta_1 \cdots \sin\theta_{n-1} \,.
\end{aligned}$$

(Spherical co-ordinates in n dimensions. The θ values run from 0 to π except last θ from 0 to 2π.) Derivative formulas:

$$\frac{\partial Z_i}{\partial U} = \frac{1}{2U} Z_i$$

and

$$\frac{\partial Z_i}{\partial \theta_j} = \begin{cases} 0 & j > i \\ -Z_i \tan\theta_i & j = i \\ Z_i \cot\theta_j & j < i \,. \end{cases}$$

Fix $n = 3$ to clarify the formulas. Use shorthand $R = \sqrt{U}$.

Matrix of partial derivatives is

$$\begin{bmatrix} \frac{\cos\theta_1}{2R} & -R\sin\theta_1 & 0 \\ \frac{\sin\theta_1\cos\theta_2}{2R} & R\cos\theta_1\cos\theta_2 & -R\sin\theta_1\sin\theta_2 \\ \frac{\sin\theta_1\sin\theta_2}{2R} & R\cos\theta_1\sin\theta_2 & R\sin\theta_1\cos\theta_2 \end{bmatrix}.$$

Find determinant:

$$U^{1/2}\sin(\theta_1)/2$$

(non-negative for all U and θ_1).

General n: every term in the first column contains a factor $U^{-1/2}/2$ while every other entry has a factor $U^{1/2}$.

FACT: multiplying a column in a matrix by c multiplies the determinant by c.

SO: Jacobian of transformation is

$$u^{(n-2)/2}u^{-1/2}/2 \times h(\theta_1, \theta_{n-1})$$

for some function, h, which depends only on the angles.

Thus joint density of $U, \theta_1, \ldots \theta_{n-1}$ is

$$(2\pi)^{-n/2} \exp(-u/2) u^{(n-2)/2} h(\theta_1, \cdots, \theta_{n-1})/2 \,.$$

To compute the density of U we must do an $n-1$ dimensional multiple integral $d\theta_{n-1} \cdots d\theta_1$.

Answer has the form

$$c u^{(n-2)/2} \exp(-u/2)$$

for some c.

Evaluate c by making

$$\int f_U(u)du = c\int_0^\infty u^{(n-2)/2}\exp(-u/2)du$$
$$= 1.$$

Substitute $y = u/2$, $du = 2dy$ to see that

$$c2^{n/2}\int_0^\infty y^{(n-2)/2}e^{-y}dy = c2^{n/2}\Gamma(n/2)$$
$$= 1.$$

CONCLUSION: the χ_n^2 density is

$$\frac{1}{2\Gamma(n/2)}\left(\frac{u}{2}\right)^{(n-2)/2} e^{-u/2}1(u > 0)\,.$$

Fourth part: consequence of first 3 parts and def'n of t_ν distribution.

Defn: $T \sim t_\nu$ if T has same distribution as

$$Z/\sqrt{U/\nu}$$

for $Z \sim N(0,1)$, $U \sim \chi^2_\nu$ and Z, U independent.

Derive density of T in this definition:

$$\begin{aligned} P(T \le t) &= P(Z \le t\sqrt{U/\nu}) \\ &= \int_0^\infty \int_{-\infty}^{t\sqrt{u/\nu}} f_Z(z) f_U(u) dz du \end{aligned}$$

Differentiate wrt t by differentiating inner integral:

$$\frac{\partial}{\partial t} \int_{at}^{bt} f(x) dx = b f(bt) - a f(at)$$

by fundamental thm of calculus. Hence

$$\frac{d}{dt} P(T \le t) = \int_0^\infty \frac{f_U(u)}{\sqrt{2\pi}} \left(\frac{u}{\nu}\right)^{1/2} \exp\left(-\frac{t^2 u}{2\nu}\right) du \, .$$

Plug in

$$f_U(u) = \frac{1}{2\Gamma(\nu/2)}(u/2)^{(\nu-2)/2}e^{-u/2}$$

to get

$$f_T(t) = \frac{\int_0^\infty (u/2)^{(\nu-1)/2}e^{-u(1+t^2/\nu)/2}du}{2\sqrt{\pi\nu}\Gamma(\nu/2)}.$$

Substitute $y = u(1+t^2/\nu)/2$, to get

$$dy = (1+t^2/\nu)du/2$$

$$(u/2)^{(\nu-1)/2} = [y/(1+t^2/\nu)]^{(\nu-1)/2}$$

leading to

$$f_T(t) = \frac{(1+t^2/\nu)^{-(\nu+1)/2}}{\sqrt{\pi\nu}\Gamma(\nu/2)}\int_0^\infty y^{(\nu-1)/2}e^{-y}dy$$

or

$$f_T(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}.$$

Multivariate Normal samples: Distribution Theory

Theorem: Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variables. Then

1. $\bar{\mathbf{X}}$ (sample mean)and $\mathbf{S}$ (sample variance-covariance matrix) are independent.

2. $n^{1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim MVN(0, \mathbf{I})$.

3. $(n-1)\mathbf{S} \sim \text{Wishart}_p(n-1, \boldsymbol{\Sigma})$.

4. $T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$ is Hotelling's T^2. $(n-p)T^2/(p(n-1))$ has an $F_{p,n-p}$ distribution.

Proof: Let $\mathbf{X}_i = \mathbf{A}\mathbf{Z}_i + \boldsymbol{\mu}$ where $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$ and $\mathbf{Z}_1, \ldots, \mathbf{Z}_p$ are independent $MVN(0, \mathbf{I})$.

So $\mathbf{Z} = (\mathbf{Z}_1^T, \ldots, \mathbf{Z}_p^T)^T \sim MVN_p(0, \mathbf{I})$.

Note that $\bar{\mathbf{X}} = \mathbf{A}\bar{\mathbf{Z}} + \boldsymbol{\mu}$ and

$$\begin{aligned}(n-1)\mathbf{S} &= \sum (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \\ &= \mathbf{A} \sum (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T \mathbf{A}^T\end{aligned}$$

Thus

$$n^{1/2}(\bar{X} - \mu) = \mathbf{A} n^{1/2} \bar{\mathbf{Z}}$$

and

$$T^2 = \left(n^{1/2}\bar{\mathbf{Z}}\right)^T \mathbf{S}_{\mathbf{Z}}^{-1} \left(n^{1/2}\bar{\mathbf{Z}}\right)$$

where

$$\mathbf{S}_Z = \sum (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T / (n-1).$$

Consequences. In 1, 2 and 4: can assume $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} = \mathbf{I}$. In 3 can take $\boldsymbol{\mu} = 0$.

Step 1: Do general $\mathbf{\Sigma}$. Define

$$\mathbf{Y} = (\sqrt{n}\bar{\mathbf{Z}}^T, \mathbf{Z}_1^T - \bar{\mathbf{Z}}^T, \ldots, \mathbf{Z}_n^T - \bar{\mathbf{Z}}^T)^T .$$

(So $\mathbf{Y}$ has dimension $p(n+1)$.) Clearly $\mathbf{Y}$ is MVN with mean 0.

Compute variance covariance matrix

$$\left[\begin{array}{cc} \mathbf{I}_{p\times p} & 0 \\ 0 & \mathbf{Q}^* \end{array}\right]$$

where $\mathbf{Q}^*$ has a pattern. It is a $p \times p$ patterned matrix with entry ij being

$$\begin{aligned} \text{Cov}(\mathbf{Z}_i - \bar{\mathbf{Z}}, \mathbf{Z}_j - \bar{\mathbf{Z}}) &= \begin{cases} -\mathbf{\Sigma}/n & i \neq j \\ (n-1)\mathbf{\Sigma}/n & i = j \end{cases} \\ &= \mathbf{Q}_{ij}\mathbf{\Sigma} \end{aligned}$$

Kronecker Products

Defn: If $\mathbf{A}$ is $p \times q$ and $\mathbf{B}$ is $r \times s$ then $\mathbf{A} \bigotimes \mathbf{B}$ is the $pr \times qs$ matrix with the pattern

$$\begin{bmatrix} \mathbf{A}_{11}\mathbf{B} & \mathbf{A}_{12}\mathbf{B} & \cdots & \mathbf{A}_{1q}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{A}_{p1}\mathbf{B} & \mathbf{A}_{p2}\mathbf{B} & \cdots & \mathbf{A}_{pq}\mathbf{B} \end{bmatrix}$$

So our variance covariance matrix is

$$\mathbf{Q}^* = \mathbf{Q} \bigotimes \mathbf{\Sigma}$$

Conclusions so far:

1) $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent.

2) $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim MVN(0, \mathbf{\Sigma})$

Next: Wishart law.

Defn: The Wishart$_p(n,\mathbf{\Sigma})$ distribution is the distribution of

$$\sum_1^n \mathbf{Z}_i\mathbf{Z}_i^T$$

where $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$ are iid $MVN_p(0,\mathbf{\Sigma})$.

Properties of Wishart.

1) If $\mathbf{A}\mathbf{A}^t = \mathbf{\Sigma}$ then

$$\text{Wishart}_p(0,\mathbf{\Sigma}) = \mathbf{A}\text{Wishart}_p(0,\mathbf{I})\mathbf{A}^T$$

2) if $\mathbf{W}_i, i = 1,2$ independent Wishart$_p(n_i,\mathbf{\Sigma})$ then

$$\mathbf{W}_1 + \mathbf{W}_2 \sim \text{Wishart}_p(n_1+n_2,\mathbf{\Sigma}).$$

Proof of part 3: rewrite

$$\sum(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T$$

in form

$$\sum_{j=1}^{n-1} \mathbf{U}_i \mathbf{U}_i^T$$

for $\mathbf{U}_i$ iid $MVN_p(0, \mathbf{\Sigma})$. Put $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ as cols in matrix $\mathbf{Z}$ which is $p \times n$. Then check that

$$\mathbf{Z}\mathbf{Q}\mathbf{Z}^T = \sum(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T$$

Write $\mathbf{Q} = \sum \mathbf{v}_i \mathbf{v}_i^T$ for $n-1$ orthogonal unit vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$. Define

$$\mathbf{U}_i = \mathbf{Z}\mathbf{v}_i$$

and compute covariances to check that the $\mathbf{U}_i$ are iid $MVN_p(0, \mathbf{\Sigma})$. Then check that

$$\mathbf{Z}\mathbf{Q}\mathbf{Z}^T = \sum \mathbf{U}_i \mathbf{U}_i^T$$

Proof of 4: suffices to have $\mathbf{\Sigma} = \mathbf{I}$.

Uses further props of Wishart distribution.

3: If $\mathbf{W} \sim Wishart_p(n, \mathbf{\Sigma})$ and $\mathbf{a} \in \mathbb{R}$ then

$$\frac{\mathbf{a}^T \mathbf{W} \mathbf{a}}{\mathbf{a}^T \mathbf{\Sigma} \mathbf{a}} \sim \chi^2_n$$

4: If $\mathbf{W} \sim Wishart_p(n, \mathbf{\Sigma})$ and $n \geq p$ then

$$\frac{\mathbf{a}^T \mathbf{\Sigma}^{-1} \mathbf{a}}{\mathbf{a}^T \mathbf{W}^{-1} \mathbf{a}} \sim \chi^2_{n-p+1}$$

5: If $\mathbf{W} \sim Wishart_p(n, \mathbf{\Sigma})$ then

$$\text{trace}(\mathbf{\Sigma}^{-1} \mathbf{W}) \sim \chi^2_{np}$$

6: If $\mathbf{W} \sim Wishart_{p+q}(n, \mathbf{\Sigma})$ is partitioned into components then

$$\mathbf{W}_{11} - \mathbf{W}_{12} \mathbf{W}_{22}^{-1} \mathbf{W}_{21} \sim Wishart_p(n - q, \mathbf{\Sigma}_{11.2})$$