Expectation of Quadratic forms

Let $\underline{X}_p$ be a random vector and $\underline{A}_{p\times p}$ a symmetric matrix. Then $E[\underline{X}^T \underline{A}\, \underline{X}] = \underline{\mu}^T A \underline{\mu} + tr[\underline{A}\,\underline{\Sigma}]$

Proof. $E[\underline{X}^T \underline{A}\, \underline{X}] = E[tr.[\underline{X}^T \underline{A}\, \underline{X}]] = E[tr[\underline{A}\, \underline{X}\, \underline{X}^T]]$

$= tr[\underline{A}\, E[\underline{X}\, \underline{X}^T]] = tr[\underline{A}(\underline{\Sigma} + \underline{\mu}\,\underline{\mu}^T)] = tr[\underline{A}\,\underline{\Sigma}] + tr[\underline{A}\,\underline{\mu}\,\underline{\mu}^T]$

$= tr[\underline{A}\,\underline{\Sigma}] + tr[\underline{\mu}^T \underline{A}\,\underline{\mu}] = tr[\underline{A}\,\underline{\Sigma}] + \underline{\mu}^T \underline{A}\,\underline{\mu}$

Distribution of Quadratic forms

Let $\underline{X}$ be $N_p(\underline{\mu}, \underline{\Sigma})$, $|\underline{\Sigma}| > 0$. Then $(\underline{X}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{X}-\underline{\mu}) \sim \chi^2(p)$

Proof. $\underline{Z} = \underline{\Sigma}^{-\frac{1}{2}}(\underline{X}-\underline{\mu}) \sim N_p(\underline{0}, \underline{\Sigma}^{-\frac{1}{2}}\underline{\Sigma}\,\underline{\Sigma}^{-\frac{1}{2}})$ and $\underline{\Sigma}^{-\frac{1}{2}}\underline{\Sigma}\,\underline{\Sigma}^{-\frac{1}{2}} = \underline{I}_p$

This shows that $\underline{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}$ is a vector of $N(0,1)$ variables

Further $\underline{Z}^T \underline{Z} = (\underline{X}-\underline{\mu})^T \underline{\Sigma}^{-\frac{1}{2}}\underline{\Sigma}^{-\frac{1}{2}}(\underline{X}-\underline{\mu}) = (\underline{X}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{X}-\underline{\mu})$

$= \sum_{i=1}^{p} Z_i^2$ and thereby $\chi^2(p)$

Thereby the $N_p(\underline{\mu}, \underline{\Sigma})$ distributions assigns probability $1-\alpha$ to the solid ellipsoid $\{\underline{x} : (\underline{x}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{x}-\underline{\mu}) \le \chi^2_{\alpha,p}\}$

where $\chi^2_{\alpha,p}$ is the upper (100α)th percentile of the χ^2_p distribution.

Assessing the assumption of normality

Univariate. Sampled values $x_1, \dots, x_n$ ordered from smallest to largest as $x_{(1)}, x_{(2)}, \dots, x_{(n)}$

$$P(X \le x_{(i)}) = F(x_{(i)}) \approx \frac{i-\frac{1}{2}}{n} \quad \left(\text{can use } \frac{i-3/8}{n+1/4}\right)$$

Also $F(x_{(i)}) = P(X \le x_{(i)}) = P\left(\frac{X-\mu}{\sigma} \le \frac{x_{(i)}-\mu}{\sigma}\right) = \Phi\left(\frac{x_{(i)}-\mu}{\sigma}\right)$

Thereby $\Phi^{-1}(F(x_{(i)})) \approx \Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right) \approx \frac{x_{(i)}-\mu}{\sigma}$ and a plot of $\Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$ against $x_{(i)}$ should be a straight line.

Multivariate sampled values: $\underline{x}_1, \dots, \underline{x}_n$

Let $d_i^2 = (\underline{x}_i - \bar{\underline{x}})^T \underline{S}^{-1} (\underline{x}_i - \bar{\underline{x}})$

1. Order d_i^2, $i = 1, 2, \dots, n$ from smallest to largest as $d_{(1)}^2 \le d_{(2)}^2 \le \dots \le d_{(n)}^2$
2. Graph $F^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$ against $d_{(i)}^2$, where F is the χ^2-distribution with p degrees of freedom and we should have a straight line with slope 1

Note $F^{-1}(F(d_{(i)}^2)) = d_{(i)}^2$

Therefore $F^{-1}\left(\frac{i-\frac{1}{2}}{n}\right) \approx d_{(i)}^2$

Estimation of $\underline{\mu}$ and $\underline{\Sigma}$ in a multivariate normal distribution (HS 3.3, 4.5)

Let $\underline{X}_1, \dots, \underline{X}_n$ be a p-dimensional random sample. How shall we estimate $\underline{\mu}$ and $\underline{\Sigma}$?

$$\hat{\underline{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \underline{X}_i = \underline{\bar{X}}_{p \times 1} \quad \text{and} \quad S = \frac{1}{?} \sum_{i=1}^{n} (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})^T$$

are natural estimators for $\underline{\mu}$ and $\underline{\Sigma}$

Let $\underline{X}_1, \dots, \underline{X}_n$ be independent and identical $N_p(\underline{\mu}, \underline{\Sigma})$

Then $M_{\underline{X}_i}(\underline{t}) = e^{\underline{t}^T \underline{\mu} + \frac{1}{2} \underline{t}^T \underline{\Sigma}\, \underline{t}}$, $i = 1, 2, \dots, n$

and $M_{\sum_{i=1}^{n} X_i}(\underline{t}) = \prod_{i=1}^{n} e^{\underline{t}^T \underline{\mu} + \frac{1}{2} \underline{t}^T \underline{\Sigma}\, \underline{t}} = e^{\underline{t}^T n \underline{\mu} + \frac{1}{2} \underline{t}^T n \underline{\Sigma}\, \underline{t}}$

$$\Rightarrow E\left[\sum_{i=1}^{n} X_i\right] = n\underline{\mu} \quad \text{and} \quad Cov\left[\sum_{i=1}^{n} \underline{X}_i\right] = n\underline{\Sigma}.$$

This gives $E[\underline{\bar{X}}] = \underline{\mu}$ and $Cov[\underline{\bar{X}}] = \frac{1}{n} n \underline{\Sigma} \cdot \frac{1}{n} = \frac{\underline{\Sigma}}{n}$

The maximum likelihood estimators

are $\underline{\bar{X}}$ and $S^{ML} = \frac{1}{n} \sum_{i=1}^{n} (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})^T = \frac{1}{n} \sum_{i=1}^{n} (\underline{X}_i \underline{X}_i^T - \underline{\bar{X}}\, \underline{\bar{X}}^T)$

$$\Rightarrow E[S] = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i X_i^T - n\underline{\mu}\underline{\mu}^T - n\underline{\bar{X}}\underline{\bar{X}}^T + n\underline{\mu}\underline{\mu}^T\right)$$

$$= \frac{1}{n}\Big(n\left[E(\underline{X}_i \underline{X}_i^T) - \underline{\mu}\underline{\mu}^T\right] - nE\left[\underline{\bar{X}}\underline{\bar{X}}^T - \underline{\mu}\underline{\mu}^T\right]\Big)$$

$$= Cov[\underline{X}_i] - \frac{\underline{\Sigma}}{n} = \underline{\Sigma} - \frac{\underline{\Sigma}}{n} = \frac{n-1}{n}\underline{\Sigma}$$

This shows that $\underline{S}^{ML}$ is biased,

and $\underline{S}^{ML} \cdot \frac{n}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^T = \underline{S}$

is often used as an estimator.

Idempotent matrices

~~Wel~~ Let $\underline{X} = [\underline{x}_1, \dots, \underline{x}_n]^T$

i.e $\underline{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$

Then $\bar{\underline{x}} = \frac{1}{n} \underline{X}^T \underline{1}$ where $\underline{1}^T = [1, 1, \dots, 1]_{1 \times n}$

and $\frac{1}{n} \underline{1}\,\underline{1}^T \underline{X} = \frac{1}{n} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{bmatrix} \underline{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \\ \vdots & \vdots & & \vdots \\ \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \end{bmatrix}_{n \times p}$

and $\underline{X} - \frac{1}{n} \underline{1}\,\underline{1}^T \underline{X} = (\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T)\underline{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \dots & x_{np} - \bar{x}_p \end{bmatrix}_{n \times p}$

and $\underline{S} = \dfrac{\underline{X}^T (\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T)(\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T) \underline{X}}{n-1}$

But $(\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T)(\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T) = \underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T - \frac{1}{n} \underline{1}\,\underline{1}^T + \underbrace{\frac{1}{n^2} \underline{1}\,\underline{1}^T \underline{1}\,\underline{1}^T}_{\frac{1}{n} \underline{1}\,\underline{1}^T}$

$= \underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T$

A square matrix A that satisfies $AA = A$ is called idempotent.

Example

$$A = \begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix} \quad AA = \begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix}$$

If A is idempotent and symmetric it is said to be a projection matrix $\left[I - \frac{1}{n} 1 1^T\right]$ is a projection matrix.

Rank of symmetric idempotent matrices

RM1 Let A an B be two matrices such that their product AB is defined.

Then $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$

This follows since the columns of AB are linear combinations of the columns of A and the rows in AB are linear combinations of the rows of B.

Example

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$AB = \begin{bmatrix} a_{11} b_{11} + a_{12} b_{21} & a_{11} b_{12} + a_{12} b_{22} \\ a_{21} b_{11} + a_{22} b_{21} & a_{21} b_{12} + a_{22} b_{22} \end{bmatrix}$$

$$\begin{matrix} \shortparallel & & \shortparallel \\ b_{11}\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{21}\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} & & b_{12}\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{22}\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \end{matrix}$$

We can do similar for row vectors.