Lecture 11. Multivariate Normal theory

Properties of means and covariances of vectors

- A random (column) vector $\mathbf{X} = (X_1, .., X_n)^T$ has mean

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), ..., \mathbb{E}(X_n))^T = (\mu_1, .., \mu_n)^T$$

  and covariance matrix

$$\text{cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = (\text{cov}(X_i, X_j))_{i,j},$$

  provided the relevant expectations exist.
- For $m \times n$ A,

$$\mathbb{E}[A\mathbf{X}] = A\boldsymbol{\mu},$$

  and

$$\text{cov}(A\mathbf{X}) = A\,\text{cov}(\mathbf{X})\,A^T, \qquad (1)$$

  since $\text{cov}(A\mathbf{X}) = \mathbb{E}\left[(AX - \mathbb{E}(AX))(AX - \mathbb{E}(AX))^T)\right] = \mathbb{E}\left[A(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T A^T\right]$.
- Define $\text{cov}(V, W)$ to be a matrix with (i, j) element $\text{cov}(V_i, W_j)$. Then $\text{cov}(A\mathbf{X}, B\mathbf{X}) = A\,\text{cov}(\mathbf{X})\,B^T$. (check. Important for later)

Multivariate normal distribution

- Recall that a univariate normal $X \sim \text{N}(\mu, \sigma^2)$ has density

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \; x \in \mathbb{R},$$

  and mgf

$$M_X(s) = \mathbb{E}[e^{sX}] = \exp\left(\mu s + \tfrac{1}{2}\sigma^2 s^2\right).$$

- $\mathbf{X}$ has a **multivariate normal distribution** if, for every $\mathbf{t} \in \mathbb{R}^n$, the rv $\mathbf{t}^T\mathbf{X}$ has a normal distribution.
- If $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{X}) = \Sigma$, we write $\mathbf{X} \sim \text{N}_n(\boldsymbol{\mu}, \Sigma)$.
- Note Σ is symmetric and is non-negative definite because by (1), $\mathbf{t}^T\Sigma\mathbf{t} = \text{var}(\mathbf{t}^T\mathbf{X}) \geq 0$.
- By (1), $\mathbf{t}^T X \sim \text{N}(\mathbf{t}^T\mu, \mathbf{t}^T\Sigma\mathbf{t})$ and so has mgf

$$M_{\mathbf{t}^T\mathbf{X}}(s) = \mathbb{E}[e^{s\mathbf{t}^T\mathbf{X}}] = \exp\left(\mathbf{t}^T\boldsymbol{\mu}s + \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}s^2\right).$$

- Hence $\mathbf{X}$ has mgf

$$M_{\mathbf{X}}(\mathbf{t}) == \mathbb{E}[e^{\mathbf{t}^T\mathbf{X}}] = M_{\mathbf{t}^T\mathbf{X}}(1) = \exp\left(\mathbf{t}^T\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right). \qquad (2)$$

Proposition 11.1

(i) *If $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ and A is $m \times n$, then $A\mathbf{X} \sim N_m(A\boldsymbol{\mu}, A\Sigma A^T)$*

(ii) *If $\mathbf{X} \sim N_n(\mathbf{0}, \sigma^2 I)$ then*

$$\frac{\|\mathbf{X}\|^2}{\sigma^2} = \frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} = \sum \frac{X_i^2}{\sigma^2} \sim \chi_n^2.$$

Proof:

(i) from exercise sheet 3.

(ii) Immediate from definition of χ_n^2. □

Note that we often write $||X||^2 \sim \sigma^2\chi_n^2$.

Proposition 11.2

Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$, where $\mathbf{X}_i$ is a $n_i \times 1$ column vector, and $n_1 + n_2 = n$. Write similarly $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where Σ_{ij} is $n_i \times n_j$. Then

(i) $\mathbf{X}_i \sim N_{n_i}(\boldsymbol{\mu}_i, \Sigma_{ii})$,

(ii) $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are independent iff* $\Sigma_{12} = 0$.

Proof:

(i) See Example sheet 3.

(ii) From (2), $M_{\mathbf{X}}(\mathbf{t}) = \exp\left(\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right)$, $\mathbf{t} \in \mathbb{R}^n$. Write

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\left(\mathbf{t}_1{}^T\boldsymbol{\mu}_1 + \mathbf{t}_2{}^T\boldsymbol{\mu}_2 + \tfrac{1}{2}\mathbf{t}_1{}^T\Sigma_{11}\mathbf{t}_1 + \tfrac{1}{2}\mathbf{t}_2{}^T\Sigma_{22}\mathbf{t}_2 + \tfrac{1}{2}\mathbf{t}_1{}^T\Sigma_{12}\mathbf{t}_2 + \tfrac{1}{2}\mathbf{t}_2{}^T\Sigma_{21}\mathbf{t}_1\right).$$

From (i), $M_{\mathbf{X}_i}(\mathbf{t}_i) = \exp\left(\mathbf{t}_i{}^T\boldsymbol{\mu}_i + \frac{1}{2}\mathbf{t}_i^T\Sigma_{ii}\mathbf{t}_i\right)$ so $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$, for all $\mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{pmatrix}$ iff $\Sigma_{12} = 0$.

□

Density for a multivariate normal

When Σ is positive definite, then $\mathbf{X}$ has pdf

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{|\Sigma|^{\frac{1}{2}}}\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right], \qquad \mathbf{x} \in \mathbb{R}^n.$$

Normal random samples

We now consider $\bar{X} = \frac{1}{n}\sum X_i$, and $S_{XX} = \sum(X_i - \bar{X})^2$ for univariate normal data.

Theorem 11.3

(Joint distribution of $\bar{X}$ and S_{XX}) Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n}\sum X_i$, and $S_{XX} = \sum(X_i - \bar{X})^2$. Then

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$;

(ii) $S_{XX}/\sigma^2 \sim \chi^2_{n-1}$;

(iii) $\bar{X}$ *and* S_{XX} *are independent.*

Proof

We can write the joint density as $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \sigma^2 I)$, where $\boldsymbol{\mu} = \mu\mathbf{1}$ ($\mathbf{1}$ is a $n \times 1$ column vector of 1's).

Let A be the $n \times n$ orthogonal matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\times1}} & \frac{-1}{\sqrt{2\times1}} & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\times2}} & \frac{1}{\sqrt{3\times2}} & \frac{-2}{\sqrt{3\times2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{bmatrix}.$$

So $A^TA = AA^T = I$. (check)

(Note that the rows form an orthonormal basis of $\mathbb{R}^n$.)

(Strictly, we just need an orthogonal matrix with the first row matching that of A above.)

- By Proposition 11.1(i), $\mathbf{Y} = A\mathbf{X} \sim \mathrm{N}_n(A\boldsymbol{\mu}, A\sigma^2 I A^T) \sim \mathrm{N}_n(A\boldsymbol{\mu}, \sigma^2 I)$, since $AA^T = I$.
- We have $A\boldsymbol{\mu} = \begin{pmatrix} \sqrt{n}\mu \\ 0 \\ . \\ . \\ 0 \end{pmatrix}$, so $Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n}\bar{X} \sim \mathrm{N}(\sqrt{n}\mu, \sigma^2)$ (Prop 11.1 (ii))
  and $Y_i \sim \mathrm{N}(0, \sigma^2), i = 2, ..., n$ and $Y_1, ..., Y_n$ are independent.
- Note also that

$$\begin{aligned} Y_2^2 + \ldots + Y_n^2 &= \mathbf{Y}^T\mathbf{Y} - Y_1^2 = \mathbf{X}^T A^T A \mathbf{X} - Y_1^2 = \mathbf{X}^T\mathbf{X} - n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = S_{XX}. \end{aligned}$$

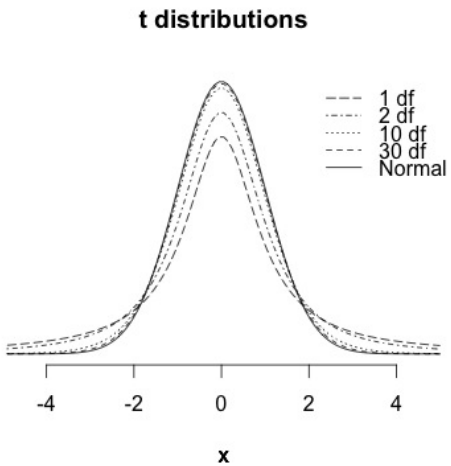
- To prove (ii), note that $S_{XX} = Y_2^2 + \ldots + Y_n^2 \sim \sigma^2 \chi^2_{n-1}$ (from definition of χ^2_{n-1}).
- FInally, for (iii), since Y_1 and $Y_2, ..., Y_n$ are independent (Prop 11.2 (ii)), so are $\bar{X}$ and S_{XX}. $\square$

Student's t-distribution

- Suppose that Z and Y are independent, $Z \sim N(0, 1)$ and $Y \sim \chi^2_k$.
- Then $T = \frac{Z}{\sqrt{Y/k}}$ is said to have a t-**distribution** on k degrees of freedom, and we write $T \sim t_k$.
- The density of t_k turns out to be

$$f_T(t) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\sqrt{\pi k}} \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}, \qquad t \in \mathbb{R}.$$

- This density is symmetric, bell-shaped, and has a maximum at $t = 0$, rather like the standard normal density.
- However, it can be shown that $\mathbb{P}(T > t) > \mathbb{P}(Z > t)$ for all $t > 0$, and that the t_k distribution approaches a normal distribution as $k \to \infty$.
- $\mathbb{E}_k(T) = 0$ for $k > 1$, otherwise undefined.
- $\mathrm{var}_k(T) = \frac{k}{k-2}$ for $k > 2$, $= \infty$ if $k = 2$, otherwise undefined.
- $k = 1$ is known as the Cauchy distribution, and has an undefined mean and variance.



Let $t_k(\alpha)$ be the upper $100\alpha\%$ point of the t_k- distribution, so that $\mathbb{P}(T > t_k(\alpha)) = \alpha$. There are tables of these percentage points.

Application of Student's t-distribution to normal random samples

- Let $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$.
- From Theorem 11.3 $\bar{X} \sim N(\mu, \sigma^2/n)$ so $Z = \sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1)$.
- Also $S_{XX}/\sigma^2 \sim \chi^2_{n-1}$ independently of $\bar{X}$ and hence of Z.
- Hence

$$\frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{S_{XX}/((n-1)\sigma^2)}} \sim t_{n-1}, \text{ ie } \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S_{XX}/(n-1)}} \sim t_{n-1}. \qquad (3)$$

- Let $\tilde{\sigma}^2 = \frac{S_{XX}}{n-1}$. Note this is an unbiased estimator, as $\mathbb{E}(S_{XX}) = (n-1)\sigma^2$.
- Then a $100(1-\alpha)\%$ CI for μ is found from

$$1 - \alpha = \mathbb{P}\left(-t_{n-1}(\tfrac{\alpha}{2}) \le \frac{\sqrt{n}(\bar{X} - \mu)}{\tilde{\sigma}} \le t_{n-1}(\tfrac{\alpha}{2})\right)$$

  and has endpoints

$$\bar{X} \pm \frac{\tilde{\sigma}}{\sqrt{n}} t_{n-1}(\tfrac{\alpha}{2}).$$

- See example sheet 3 for use of t distributions in hypothesis tests.