Lecture 1: Review of the Multivariate Normal Distribution

Why study normal/Gaussian distribution

- Diffusion and Brownian motion
- Distribution of measurment errors
- Animal characteristics
- Central Limit Theorem

Example: Measurments of iris flowers

For each flower we record the following

$$X_i = \begin{bmatrix} X_{i,se} \\ X_{i,sw} \\ X_{i,pl} \\ X_{i,pw} \end{bmatrix} \begin{matrix} \to \text{sepal length} \\ \to \text{sepal width} \\ \to \text{petal length} \\ \to \text{petal width} \end{matrix}$$

R code : iris_data_visualization.R

Multivariate Normal (Gaussian) Distribution

Definition: A random vector $X = (X_1, \dots, X_k)^T$ follows a multivariate normal distribution if its probability density function is

$$f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right\}$$

- mean vector $\mu = E(X) \in \mathbb{R}^k$
- covariance matrix $\Sigma = E\left[(X-\mu)(X-\mu)^T\right]$

  must be symmetric and positive definite

Recall that the probability density function (pdf) uniquely defines the distribution of a random variable

Geometry of Gaussian Distribution

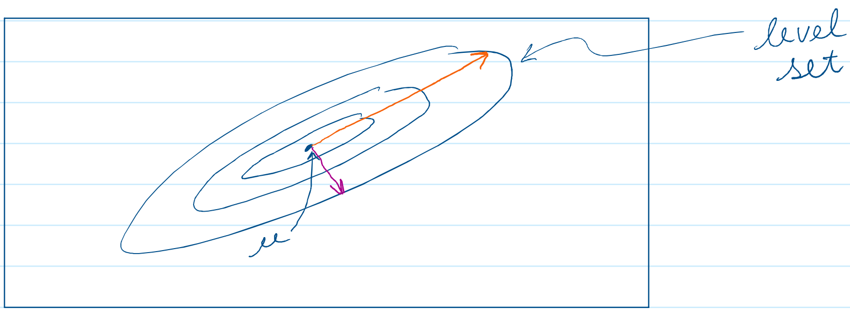
Mahalanobis distance defines the geometry

$$D_M(X) = \sqrt{(X-\mu)^T \Sigma^{-1} (X-\mu)}.$$

$$D_M(X) = \sqrt{(X-\mu)^T \Sigma^{-1} (X-\mu)}$$

Surfaces of constant density (level sets) are hyper-ellipsoids

R code : bivariate_normal_level_sets.R



Eigen-decomposition : $\Sigma = V \Lambda V^T$

- Eigenvectors in V define the principal axes (directions of greatest variation)
- Eigenvalues λ_i define the variance along those axes

Linear Transformations

<u>Theorem</u>: If $X \sim N_k(\mu, \Sigma)$, then for any constant matrix $A_{m\times k}$ and vector $b_{m \times 1}$

$$Y = AX + b \Rightarrow Y \sim N_k(A\mu + b, A\Sigma A^T)$$

R code: bivariate_normal_sampling.R

Consequences

- Marginalization: If X is MVN, any subset of its components is multivariate normal
- Standardization : Let $\Sigma^{1/2}$ be a decomposition $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T$. If $Z \sim N(0, I)$ and $X = \mu + \Sigma^{1/2} Z$, then $X \sim N(\mu, \Sigma)$
  Conversely, if $X \sim N(\mu, \Sigma)$ then

$$\Sigma^{-1/2}(X-\mu) \sim N(0, I)$$

Partitioning and conditioning

Partitioning and conditioning

If we partition X into subvectors as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

the conditional distribution $X_1 \mid X_2 = x_2 \sim N_k(\mu_{1|2}, \Sigma_{1|2})$

where conditional mean is defined as

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu)$$

conditional covariance

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \leftarrow \text{Schur Complement}$$

Moment generating functions

Definition: Let $X \sim N_k(\mu, \Sigma)$ be a k-dimentional multivariate normal random variable. The moment generating function (mgf) of X is defined for any vector $t \in \mathbb{R}^k$ as

$$\psi_X(t) = E\left[e^{t^T X}\right] = \exp\left\{t^T\mu + \frac{1}{2}t^T\Sigma t\right\}$$

The mgf uniquely defines a distribution

Inference from Data

Suppose that we observe data as realizations of n independent MVN random variables

$$Y_1, \ldots, Y_n \overset{ind}{\sim} N_k(\mu, \Sigma)$$

μ, Σ are population quantities

Definition: We will refer to the joint density $f(y_1, \ldots, y_n \mid \mu, \Sigma)$ of the data as the likelihood function.

Definition : a prior distribution models subjective likelihood function.

<u>Definition</u> : a prior distribution models subjective uncertainty about a population quantity using a probability distribution

<u>Example</u> : although μ is unknown, we may know that it should take values over the entire real line, and is most likely to lie around some value μ_0. Thus, a reasonable prior distribution for μ might be

$$\mu \sim N_k(\mu_0, \Sigma_0)$$

<u>Definition</u> : The posterior distribution has density

$$f(\mu \mid y_1,..,y_n) = \frac{f(y_1,\dots,y_n \mid \mu)\, f(\mu)}{\int f(y_1,\dots,y_1 \mid \mu)\, f(\mu)\, d\mu}$$

$$\propto f(y_1,..,y_1 \mid \mu)\, f(\mu)$$

<u>Example</u>: Let's return to our example with unknown μ. The posterior density of μ is:

$$\mu \mid Y_1 = y_1, \dots, Y_n = y_n \sim N_k(\mu_n, \Sigma_n)$$

$$\Sigma_n^{-1} = \Sigma_0^{-1} + n\Sigma^{-1}$$

$$\mu_n = \Sigma_n(\Sigma_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y})$$

prior mean

sample mean of $y_1,..,y_n$

$$\bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

R code: normal_updating_example.R