Parameter estimation for multivariate distributions

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke
rihuan.ke@bristol.ac.uk

Teaching Block 1, 2024

University of BRISTOL

What we will cover today

We will introduce the concept of a random vector

We will introduce the family of multivariate Gaussian distributions.

We will also consider parameter estimation for multivariate Gaussian distributions

Multivariate distributions

Multivariate distributions

We often need to think about distributions involving multiple features.

```
## # A tibble: 9 x 8
## # Groups:   species [3]
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>           <dbl>         <dbl>            <int>       <int> <fct>
## 1 Adelie  Dream            37.3          16.8              192        3000 fema~
## 2 Adelie  Torge~           33.5          19                190        3600 fema~
## 3 Adelie  Biscoe           45.6          20.3              191        4600 male
## 4 Chinst~ Dream            49.6          18.2              193        3775 male
## 5 Chinst~ Dream            58            17.8              181        3700 fema~
## 6 Chinst~ Dream            52.7          19.8              197        3725 male
## 7 Gentoo  Biscoe           49.6          15                216        4750 male
## 8 Gentoo  Biscoe           43.6          13.9              217        4900 fema~
## 9 Gentoo  Biscoe           49.5          16.1              224        5650 male
## # ... with 1 more variable: year <int>
```

n examples

d variables

To model the relationships between these features we must consider multivariate distributions

Univariate analysis is about a single variable, e.g., bill length

Multivariate analysis considers several variables, e.g., (bill length, bill depth, body mass)

Random variables and random vectors

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. A random variable is a mapping $X : \Omega \to \mathbb{R}$, such that for every $a, b \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \in [a, b]\}$ is an event in $\mathcal{E}$

Example: rolling a dice, we model its outcome with a random variable X (the value of X is a single number taken from $\{1, 2, \cdots, 6\}$).

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. A random vector is a mapping $X : \Omega \to \mathbb{R}^d$, such that for every $a_1, \cdots, a_d, b_1, \cdots, b_d \in \mathbb{R}$ with each $a_i \leq b_i$, we have $\left\{\omega \in \Omega : X(\omega) \in \prod_{i=1}^{d} [a_i, b_i]\right\}$ is an event in $\mathcal{E}$

Note: here the mapping X outputs a vector (with d elements): $X(\omega) := (X_1(\omega), X_2(\omega), \cdots, X_d(\omega))$.

Example: rolling two dice, we model their outcome with a random vector X (the value of X is a pair of numbers e.g. (1,2), (3,3), $\cdots$).

Probability density function

Continuous random variables are specified by a probability density function $f_X : \mathbb{R} \to [0, \infty)$ with $\int_{-\infty}^{\infty} f_X(x) = 1$. For all $a, b \in \mathbb{R}$, we have

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$$

Continuous random vectors $X := (X_1, X_2, \cdots, X_d)$ are specified by a multivariate probability density function $f_X : \mathbb{R}^d \to [0, \infty)$ with $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \cdots, x_d)dx_d \cdots dx_1 = 1$. For all $a_1, \cdots, a_d, b_1, \cdots, b_d \in \mathbb{R}$ with each $a_i \leq a_d$, we have

$$\mathbb{P}(X \in [a_1, b_1] \times \cdots \times [a_d, b_d]) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f_X(x_1, \cdots, x_d)dx_d \cdots dx_1$$

note: here $[a_1, b_1] \times \cdots \times [a_d, b_d]$ is a Cartesian product defined as follow:

$$[a_1, b_1] \times \cdots \times [a_d, b_d] := \{(x_1, \cdots, x_d) : x_1 \in [a_1, b_1], \cdots, x_d \in [a_d, b_d]\ \}$$

Gaussian random variables

A classical example of continuous random variables is a Gaussian.

A Gaussian random variable is a continuous random variable X with the probability density function $f_{\mu,\sigma} : \mathbb{R} \to [0, \infty)$ defined by

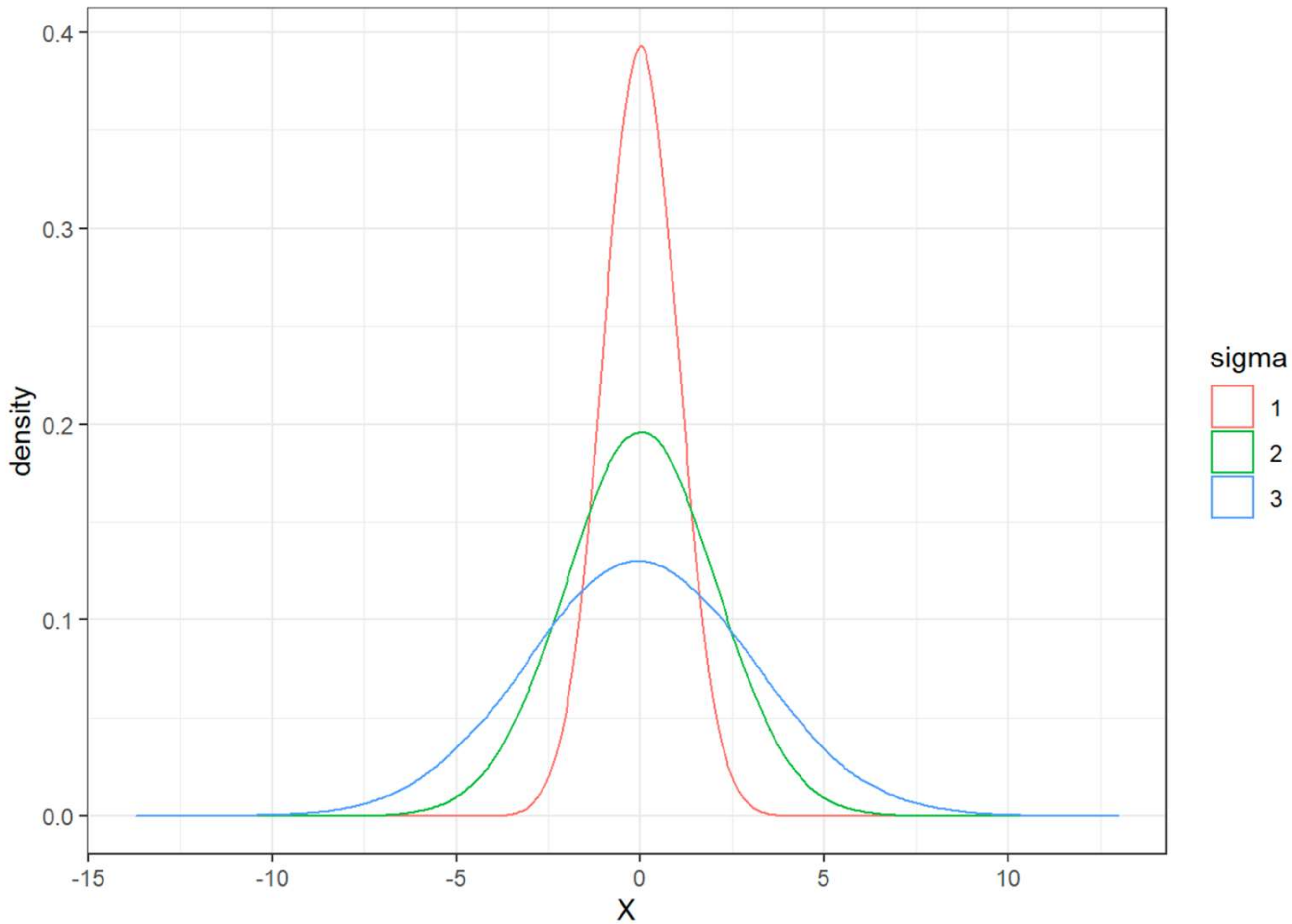
$$f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \text{ for all } x \in \mathbb{R}.$$

A Gaussian random variable is often referred to as a normal random variable.

We often write $X \sim \mathcal{N}(\mu, \sigma^2)$ to mean X is Gaussian with parameters μ, σ.

For a Gaussian random variable X, we have $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

Gaussian random variables



$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

For a Gaussian random variable X, we have $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

Multivariate Gaussians

A classical example of continuous random vector is a multivariate Gaussian $X = (X_1, \cdots, X_d)$.

Its parameters are

1) A mean vector: $\mu = \mathbb{E}(X) \in \mathbb{R}^d$
2) A covariance matrix: $\Sigma = \mathbb{E}\big[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T\big] \in \mathbb{R}^{d\times d}$.

The probability density function $f_{\mu,\Sigma} : \mathbb{R}^d \to (0, \infty)$ is given by

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

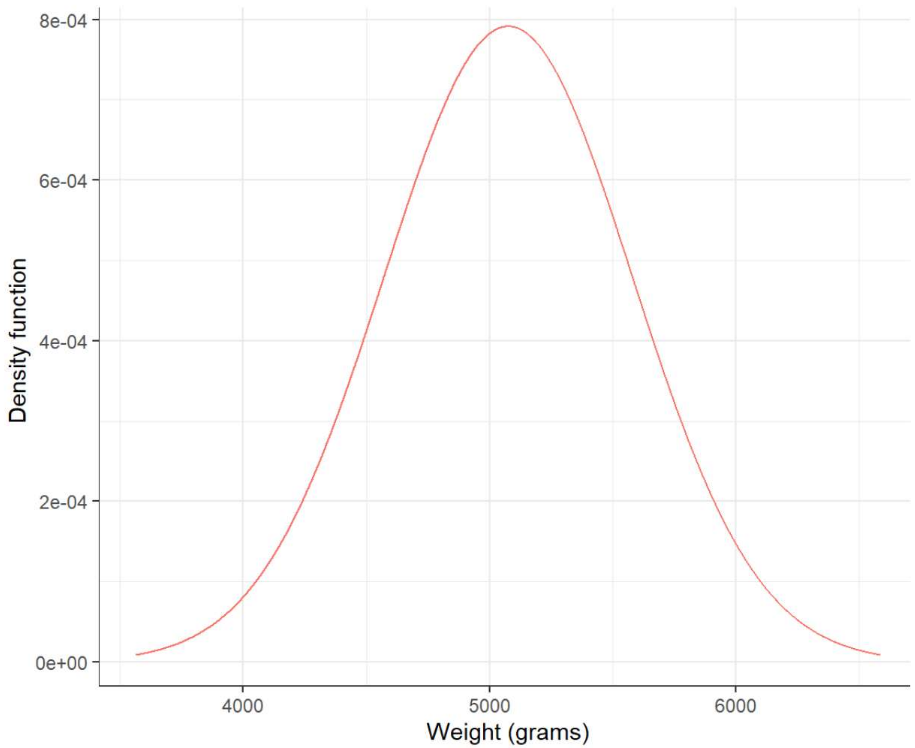
where $x \in \mathbb{R}^d$, $|\Sigma|$ is the determinant of Σ, Σ^{-1} is the inverse of Σ, and $(x-\mu)^T$ is the transpose of $x - \mu$.

For comparison: $f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$.

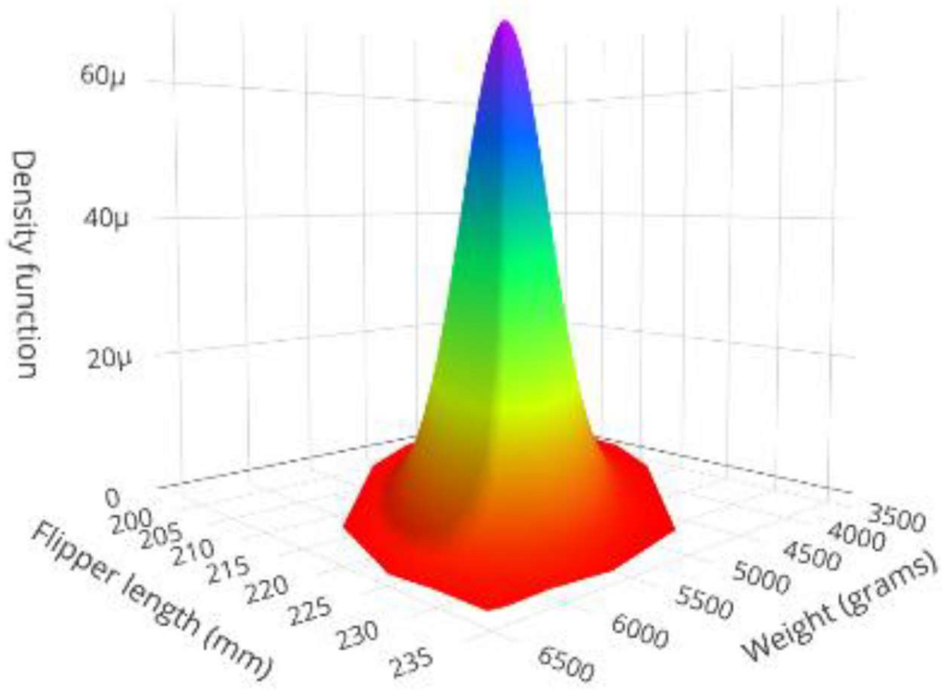
Note: if $d = 1$, then the multivariate Gaussian defined above reduces to a univariate Gaussian that we discussed previously.

Multivariate Gaussians

A univariate Gaussian and a bivariate Gaussian



univariate Gaussian



bivariate Gaussian

Parameter estimation for multivariate Gaussians

Parameter estimation for multivariate Gaussians

Suppose that we have a i.i.d. sample $X_1, X_2, \cdots, X_n \sim \mathcal{N}(\mu, \Sigma)$ from a multivariate Gaussian distribution.

The population parameters are $\mu := \mathbb{E}(X)$ and $\Sigma := \mathbb{E}\big[(X - E(X)(X - E(X))^T\big]$.

<u>Parameter estimation</u>: Given the sample $X_1, \cdots, X_n$, we want to estimate the population parameter μ and Σ.

1. The sample mean $\overline{X}$ is both the MVUE and the MLE for μ.

2. $\hat{\Sigma}_U = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})^T \in \mathbb{R}^{d \times d}$ is the MVUE for $\Sigma \in \mathbb{R}^{d \times d}$.

3. $\hat{\Sigma}_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})^T \in \mathbb{R}^{d \times d}$ is the MLE for $\Sigma \in \mathbb{R}^{d \times d}$.

Example

Example. fit a multivariate model for our Gentoo penguins.

Our sample:

```
penguins_gwf<-penguins%>%
  filter(species=="Gentoo")%>%
  select(body_mass_g,flipper_length_mm)
```

We model the body mass and flipper length with a bivariate Gaussian random vector.

we want to fit the model to the data. To do this, we obtain estimates of the population mean and population covariance

Example

MLE for population mean:

```
mu_gwf<-map_dbl(penguins_gwf,~mean(.x,na.rm=1)) # MLE estimate of the mean
mu_gwf
```

```
##       body_mass_g flipper_length_mm
##          5076.016           217.187
```

2. $\hat{\Sigma}_U = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T \in \mathbb{R}^{d\times d}$

MVUE estimate of the covariance (matrix):

```
Sigma_gwf<-cov(penguins_gwf,use="complete.obs") # MVUE estimate of the covariance
Sigma_gwf
```

```
##                   body_mass_g flipper_length_mm
## body_mass_g        254133.180        2297.14448
## flipper_length_mm    2297.144          42.05491
```

3. $\hat{\Sigma}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T \in \mathbb{R}^{d\times d}.$

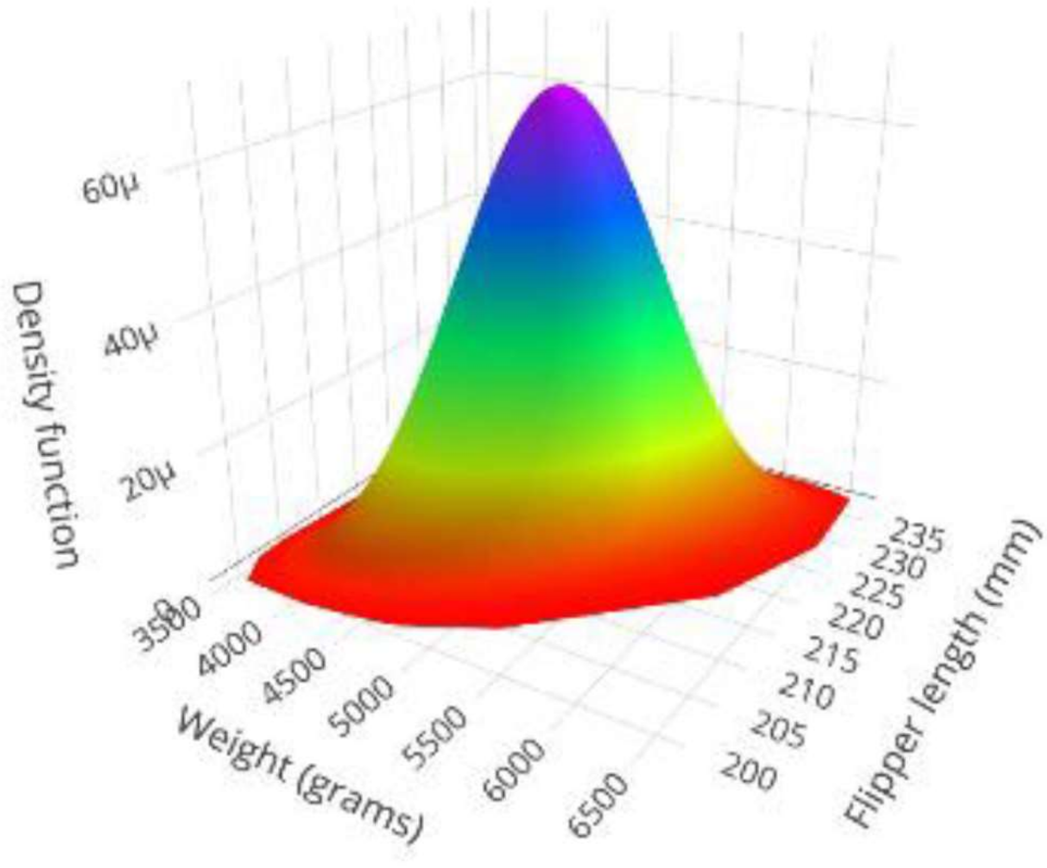
MLE estimate of the covariance (matrix):

```
Sigma_gwf_MLE<-cov(penguins_gwf,use="complete.obs")*(n-1)/n # MLE estimate of the covariance
Sigma_gwf_MLE
```

```
##                   body_mass_g flipper_length_mm
## body_mass_g        252083.719        2278.61912
## flipper_length_mm    2278.619          41.71576
```

Example

The density function of the fitted model (i.e., the bivariate Gassuain with the estimated mean and covariance):



What have we covered?

We introduced the concept of a random vector.

We saw that continuous random vectors can be understood via probability density functions.

We introduced the concept of a multivariate Gaussian distribution.

We also considered parameter estimation for multivariate Gaussian distributions.

Thanks for listening!

Dr. Rihuan Ke
rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science