43075-01 Probabilistic Shape Modelling

Lecturers
Dr. Marcel Lüthi ⟨marcel.luethi@unibas.ch⟩

Introduction	26. April 2022
Discussion	03. Mai 2022

Exercise 6 — Stature and sex prediction

Introduction

Our final goal in the project is to estimate the stature and sex, given only a fragment of a bone. We will do this in two steps: In the first step we will compute a distribution of plausible reconstructions of the full shape from the fragment. From these reconstructions we can easily obtain measurements such as the length of the femur and the distance between the top of the greater trochanter and the lesser trochanter, which are known to be good predicters of stature and sex of the person (see Figure 1 for an illustration of the anatomy of the femoral head).
In this exercise we start with the second step. We develop methods to predict the stature of the person from the length of the femur and its sex from the trochanter distance.

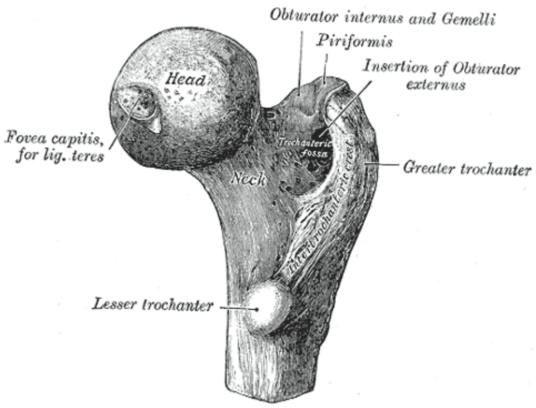

Figure 1: Anatomy of the femoral head. Source: By Henry Vandyke Carter - Henry Gray (1918) Anatomy of the Human Body

6.1 Preparation

Theory

Work through Week 1 and 2 of the probabilistic fitting online course. In particular, study carefully how Bayesian linear regression works. Study the implementation of a basic Metropolis-Hastings sampler given in Tutorial 14.

Data

Download the file `StatureAndSex.csv` from the Adam workspace:
`https://adam.unibas.ch/goto_adam_file_1396380.html`.

This file contains measurements for 43 subjects. The first field is the id, and it is used to identify the subject. The id is followed by the sex, the stature (in mm), measurements of the femur length and a measurement of the trochanter distance.
To read the CVS data, you can use the following procedure:

```
case class Measurements(
    id: String,
    sex: String,
    stature: Double,
    lengthOfBone: Double,
    trochanterDistance: Double
)

def readData(csvFile: java.io.File): Seq[Measurements] = {
    val lines = scala.io.Source.fromFile(csvFile).getLines.toList
    val header = lines.head.split(",")
    lines.tail.map(line => {
        val values = line.split(",")
        Measurements(
            values(0),
            values(1),
            values(2).toDouble,
            values(3).toDouble,
            values(4).toDouble
        )
    })
}
```

Software

The upcoming release of Scalismo (currently Version 0.91-RC3) introduces functionality that makes developing and debugging Metropolis-Hastings samplers easier. While it is perfectly possible to continue using Scalismo version 0.90, you may want to give the new version a try for this second part of the project. To follow the new version, you need to change the version of the scalismo tutorials using the switch-box on the upper-right corner of the tutorial pate to next *as shown in Figure 2. In your source files, you will need to specify the new version of Scalismo in the header of your source files:*

```
//> using scala "2.13"
//> using lib "ch.unibas.cs.gravis::scalismo-ui:0.91-RC3"
```

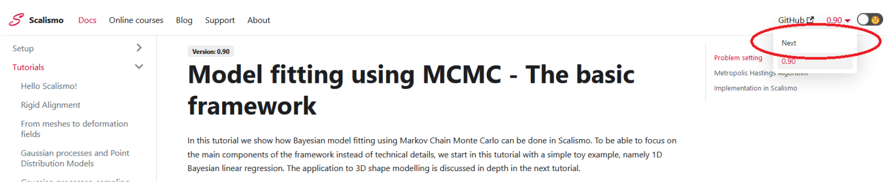

Figure 2: Choosing the scalismo-version used in the tutorials.

If you decide to switch to a the latest Scalismo version, you may want to organize your source code for this second part of the project, in a separate folder, to make sure that VSCode uses the right Scalismo version internally.

6.2. Bayesian linear regression for stature estimation

Adapt the code from tutorial 14, to train a linear regression, which relates the measured femur length and the stature of the person. Implement a method, which predicts the stature of a person for a given length.
Think about the following questions:

- What are good priors for slope, intercept and variance?
- What are good parameters to use to define the proposal distributions?
- What is the acceptance rate of your proposals? Is it what you would expect?
- What are the posterior values for the parameters slope, intercept and variance for your data? What is the uncertainty of the parameters you obtain as a result?
  - Is it justified to use a point-estimate (such as the MAP or expected value) for these parameters for predicting the stature from the length, or should you model the uncertainty for the predictions? How would you include the uncertainty?

6.3. Logistic regression for sex prediction

To predict the sex, we are using a technique called logistic regression. As sex, in contrast to stature, is a categorical variable with only two possible outcomes, we cannot use the simple linear relationship $y \sim ax+b+\epsilon$ as for predicting the stature. The idea used in logistic regression is, to squash the outcome of the linear prediction $ax+b$, using the sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$. The resulting value $\sigma(ax+b)$ has values in $[0,1]$. We assume that the resulting random variable follows a Bernoulli distribution. This results in the following model:

$$y \sim \text{Bernoulli}(\sigma(a \cdot x + b)).$$

For given pairs of observations $\{(x_i, y_i)\}_{i=1}^n$, the likelihood function becomes

$$\prod_{i=1}^n p(y_i|a,b,x_i) = \prod_{i=1}^n y \sim \text{Bernoulli}(\sigma(a \cdot x_i + b)).$$

Adapt the code you wrote for Bayesian linear regression, by changing the likelihood function in the evaluator accordingly. Implement a method, which predicts the sex of a person for a given trochtanter distance.

- What are good priors for slope and intercept?
- What are good parameters to use to define the proposal distributions?
- What is the acceptance rate of your proposals? Is it what you would expect?
- What are the posterior values for the parameters slope and intercept ? What is the uncertainty of the parameters you obtain as a result?
  - Is it justified to use a point-estimate (such as the MAP or expected value) for these parameters for predicting the stature from the length, or should you model the uncertainty for the predictions? How would you include the uncertainty?

6.4. Validate the model

Draw a number of samples from the statistical shape model and measure the length of the samples using the distance between landmarks `L2` and `L5` as the length of the femur and the distance between landmarks `L2` and `L1` as the trochanter distance. Use the regression procedure above to predict the stature and sex.
Analyse the results (by computing summary statistics such as mean and standard deviation and/or by plotting the results). Critically assess your predictions. Do they correspond to what is known about the known height distribution of persons living in Switzerland? (see e.g. `https://en.wikipedia.org/wiki/Average_human_height_by_country`. Do you get as many male and female predictions as you would expect?