Statistical computing MATH10093
Coursework A 2017/18

Finn Lindgren

7/2/2018 (comments 29/2/2019)

Summary

Comments on last year's coursework: The handout–to–handin time was three weeks. Discussion of the assignment with others was permitted, but handin was of individual solutions. The work was marked out of 50, and counted for 50% of the total grade.

General credits

A total of 10 marks is awarded for general acquired skills:

1. RMarkdown used to produce the handin

2. Basic mathematical formula typesetting with RMarkdown

3. Code readability, useful code comments

The following four questions are worth 10 marks each.

Question 1: A new score

Write a new function `score_interval` that implements the Interval score:
Let L_F and U_F be lower and upper bounds of prediction intervals for some error probability α, e.g. $\alpha = 0.05$.

$$S_{\text{INT}}(F, y) = U_F - L_F + \frac{2}{\alpha}(L_F - y)\mathbb{I}(y < L_F) + \frac{2}{\alpha}(y - U_F)\mathbb{I}(y > U_F)$$

On average, this score is minimised when (L_F, U_F) is the equal tail probability interval ($\alpha/2$ on each side) under the true distribution G, and it is a proper score.

When used on output from `model_predict`, the new function should use the `lwr` and `upr` bounds. For top marks, the function should allow the user to specify α, and it should be included in the score comparisons in Question 3, with $\alpha = 1/3$ (this will require providing α for `model_predict` as well).

Question 2: The 3D printer problem

Problem description

A 3D-printer for home use typically operates by heating the end of a coil of plastic or other material just enough to squeeze through a nozzle, and onto an assembly plate, or already printed material. The object to be printed is first designed in a CAD (Computer Aided Design) program, which also calculates the expected amount of material required to print the object. When the object has been printed, we can weigh it to learn how much material was *actually*

used. Due to variable material properties, the CAD-predicted amount and the actual amount will not match perfectly. It is also expected that the variability is more noticeable for large objects. Hence, it is expected that the actual amount will depend on both the material type, identified by colour, and the nominally expected amount as calculated by the CAD program.

Question

Look at the data generating code in `CWAcode.R` (2019 lab 3: `lab03code.R`) and determine the 12-parameter model structure (ideally, describe it in mathematical notation as well as plain text).

Construct an appropriate `model_Z` function, and estimate $\boldsymbol{\theta}$ for the 3D-printer problem.

Before you start, read through Question 3, and the solutions to Lab 3!

Question 3: Assessing models for the 3D printer problem

In Lab 3, a model that used only `cad` to model the expectation and variance was used for the 3D-printer problem. Compare the new, colour aware, model with that simpler model by computing proper scores.

Briefly discuss the meaning of the results.

Hint: Since the two models need different `model_Z` functions, you may use the following technique to switch between the two models:

```
model_Z_lab <- function(...) ...
model_Z_cw <- function(...) ...
model_Z <- model_Z_lab
# ... Perform estimation/prediction for the Lab 4 model
model_Z <- model_Z_cw
# ... Perform estimation/prediction for the colour aware coursework model
```

Question 4: Finite differences (2019: will be lectured in the second half)

Choose a non-trivial function $f(\theta)$ with known derivatives, and a point θ.

Compute the error of first and second order finite differences when using an optimal h step value. For the second order derivative, also compare with the error obtained if one uses the h-value that is optimal for a first-order difference.

Refer to the notes for Lecture 2 and Lecture 3 for approximate optimality criteria, and use that when choosing h values.