Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Statistik III/Advanced Statistics
SS 2026

Exercises

Exercise 27. Reconsider the data contained in the file `orthography.txt`, which were already analyzed in a previous exercise. The data set originates from a diagnostic test of spelling difficulties in 5th and 6th graders from secondary schools in Baden-Württemberg. It includes information on the following variables:

X_1	`CFT`	Culture Fair Intelligence Test (Subtest 1)
X_2	`WM`	Phonological working memory performance (subtest of VLMT)
X_3	`sex`	Gender (dummy coded: 0 male, 1 female)
X_4	`school`	School type (dummy coded: 0 Hauptschule, 1 Realschule)
X_5	`class`	Grade level (dummy coded: 0 5th grade, 1 6th grade)
Y	`score`	Number of correct spellings in 20 words

a) In a linear regression analysis, create two models that each include either the variables `CFT`, `WM`, `sex`, or the variable `class` as predictors of spelling performance in addition to the variable `school`. Estimate the parameters of the two models using R.

b) Apply the `logLik()` function and, based on the resulting log-likelihoods, compute the respective AIC and BIC values by hand.

c) Check your results using the `AIC()` and `BIC()` functions.

d) Compare the models with respect to the AIC and BIC values. Which of the models is to be preferred?

Exercise 28. Whenever the posterior belongs to the same family of distributions as the prior then the latter is called a conjugate prior for the respective likelihood. The beta distribution is a conjugate prior for the likelihood of a binomial distribution. After observing k out of n cases, the prior beta distribution $B(\alpha, \beta)$ with parameters α, β is transformed into the posterior beta distribution $B(\alpha+k, \beta+n-k)$. Draw upon this property in solving the following problem.

Assume that in the preceding season the shooting of a biathlete in the indiviual race showed a hit rate which can be characterized by a beta distribution with parameters $\alpha = 13.5$ and $\beta = 1.5$, and expected value $\alpha/(\alpha+\beta) = 0.9$. In the first competition of the current season the biathlete scored 16 hits in 20 shots.

a) Generate graphs of the prior and the posterior based on the results of the first competition.

b) Estimate the posterior probability of a hit via the EAP estimate.

c) Obtain the MCMC approximation to the EAP estimate in Stan. (Hint: Use the Code from the slides.)

Exercise 29. A principle named after R.A. Fisher claims that a 1:1 sex ratio is an evolutionary stable strategy. Consider the following hypotheses about the probability of the birth of a boy:

$$H_1\text{: } p = 0.5$$
$$H_2\text{: } p = 0.513$$
$$H_3\text{: } p = 0.555$$

Notice that each of the hypotheses assumes a single parameter value only, so that the likelihood given this parameter value coincides with the marginal likelihood of the corresponding hypothesis. The parameter value in H_2 is close to the relative frequency observed for Germany in years 1991-2018.

Assume that in a hospital out of $n = 3296$ new born babies $m = 1715$ were boys. Based on these data, use the `dbinom()` function to compute the Bayes factors for comparing hypotheses H_1 and H_2 as well as H_1 and H_3. How do you evaluate the evidence of the data according to the classification scheme of Jeffreys (1961)?