Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Statistical Models in Computational Biology

Niko Beerenwinkel
David Dreifuss
Rudolf Schill
Auguste Rimaite

Due date: 7th Mar 2024 before 13:00

Problem 4: Responsibilities and prior for biased coins **(3 points)**

Two coins C_A and C_B have a probability of 0.7 and 0.4 for obtaining heads, respectively. We randomly choose one coin and flip it 10 times. We call this vector of flips an observation. Heads are denoted by 1 and tails by 0. The first observation is

$$D_1 = (1, 1, 0, 1, 1, 1, 0, 1, 1, 1).$$

Then, we randomly choose a coin for another 10 flips and the second observation is

$$D_2 = (0, 0, 1, 0, 0, 0, 1, 1, 0, 0).$$

The prior (mixture weights) is $P(C_1) = 0.6$ and $P(C_2) = 0.4$. Compute the responsibilities $P(C_i \mid D_j)$ of each coin for the two observations. Update the coins' mixture weights accordingly.

Hint: To compute the posterior of C_i, use the Bayes' theorem. You need to compute the likelihood of C_i given D_j and the marginal of D_j.

Problem 5: Learning a mixture model for two biased coins **(7 points)**

Two coins C_A and C_B have unknown probabilities of obtaining heads. We randomly choose a coin and flip it 100 times. We call this vector of flips an observation. Heads are denoted by 1 and tails by 0. Overall we have 200 observations. We do not know, which coin is responsible for which observation.

Compute the probability of heads for each original coin C_A and C_B. Please use the R code skeleton `CoinEM_skeleton.Rmd` on Moodle for this question.

(a) Read the data stored in the file 'coinflip.csv'. It contains a 200×100 matrix, where each observation has 100 replicated flips. (1 point)

(b) Randomly initialize the priors (mixture weights) λ_1 and λ_2 (e.g. $\lambda_1 = \lambda_2 = 0.5$), and the probabilities for heads ($= 1$) for each of the coins (e.g. from the Uniform distribution). (1 point)

(c) EM algorithm: (3 point)

- E-step: use the priors and the coin probabilities to compute the responsibilities γ, the observed log-likelihood, and the expected hidden log-likelihood.
- M-step: use the responsibilities to recompute the priors and the probability of heads (1) of each coin.
- Iterate over E- and M-step until convergence of the likelihood.

(d) Plot the Log-likelihood, the probability of heads for each coin and the mixture weights as a function of the iteration number of the algorithm, and briefly comment what you see. Plot a heatmap of the responsibilities γ at the final step. How many observations belong to each coin? (2 point)