

# Exercises Algorithmic Systems Biology

Freie Universität Berlin, SoSe 2010  
Roland Krause · Martin Vingron

## Blatt 2 · Durchführung am 4.6.2010

**Aufgabe 1** (Bayesian statistics). Suppose  $D$  and  $E$  are events with joint probability given by

	$D$	$\bar{D}$
$E$	0.10	0.05
$\bar{E}$	0.5	0.35

Find the conditional probabilities  $P(D|E)$ ,  $P(E|D)$ ,  $P(D|\bar{E})$  and  $P(E|\bar{D})$ , and the marginal probabilities  $P(D)$ ,  $P(\bar{D})$ ,  $P(\bar{E})$  and  $P(E)$ .

Are the events independent?

**Aufgabe 2** (Beta and Gamma distribution). Explore the Gamma distribution and Beta distribution by varying the parameters. Which shapes can you produce?

Plot the beta distribution using different  $\alpha$  and  $\beta$ . For the gamma distribution, it suffices to vary the shape parameter  $k$ . Why could the gamma distribution be useful in Bayesian setting? Which distribution do you get if  $k = 1$ ?

**Aufgabe 3** (Bayesian network). Each node in the gene regulatory network has two states. The connectivity is  $A \rightarrow B, A \rightarrow C, B \rightarrow D, C \rightarrow D$ . Sketch the posterior probability for each node.

What is the joint probability distribution?

What is the probability of  $D$  being in state  $d_1$ ?

What is the probability of  $P(a_1|b_1, c_1)$ ?

**Aufgabe 4** (Posterior inference). Suppose there is a Beta(4,4) prior distribution on the probability  $\theta$  that a coin will yield *head*. The coin is spun 10 times and heads appear fewer than 3 times.

You do not know how many heads were seen, only that the number is less than three. Calculate the posterior density for  $\theta$  and sketch it.

Hint: Use a binomial model to calculate the likelihood  $Pr(data|\theta)$  and the simple prior density  $p(\theta) \propto (\theta)^n(1 - \theta)^n$ .

**Aufgabe 5** (Fisher). The analysis of a gene expression experiment identifies 20 genes of the 5000 genes of yeast as significant of a particular condition. 5 genes of those genes belong to a particular GO category spanning a total of 60 genes. Is this finding significant?

How do you have to modify the result given the 3000+ GO categories?

A study reports that about 2500 genes in yeast are never widely expressed. How do you have to adjust the statistic?