Exercises for the Statistics for Bioinformaticians Lecture
Winter 2013/2014

Problem Set 1

Heike Siebert - siebert@mi.fu-berlin.de
Alena van Bömmel - mysickal@zedat.fu-berlin.de

The following problems refresh some fundamental notions from stochastics and aim at practicing the basic definitions and properties concerning Markov chains introduced in the lecture. The third problem consists of proving the remaining implications of Theorem 1.7 of the lecture.

Problem 1 (4 points)
Let $(\Omega, \mathcal{F}, P)$ be a probability space. Prove the following rules (Bayes' rules).

A With $A \in \mathcal{F}$, $P(A) > 0$, we have

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \qquad \text{(2 points).}$$

B For a partition $\{B_i \in \mathcal{F} | i \in \mathbb{N}\}$ of Ω with $P(B_i) > 0$ for all $i \in \mathbb{N}$ and $A \in \mathcal{F}$, we have

$$P(A) = \sum_{i \in \mathbb{N}} P(A|B_i) \cdot P(B_i) \qquad \text{(2 points).}$$

Problem 2 (2 points)
Let $M_1, \ldots, M_k$, $k \in \mathbb{N}$, be stochastic $n \times n$ matrices. Show that $\sum_{i=1}^{k} \alpha_i M_i$ is a stochastic matrix for all $\alpha_1, \ldots, \alpha_k \in \mathbb{R}_{\geq 0}$ with $\sum_{i=1}^{k} \alpha_i = 1$.

Problem 3 (4 points)
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(X_n)_{n \in \mathbb{N}_0}$ be a sequence of random variables on Ω taking values in $E := \{1, \ldots, l\}$. Show the implications (2)$\Rightarrow$(3) and (3)$\Rightarrow$(1) for the following statements (2 points each).

(1) $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain.

(2) There exists $\mathbb{P} \in [0,1]^{l \times l}$, $\mathbb{P} = (p_{ij})_{i,j \in \{1,\ldots,l\}}$, with $P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = p_{i_{n-1} i_n}$ for all $n \in \mathbb{N}$ and $i_0, \ldots, i_n \in E$.

(3) There exist $\mathbb{P} \in [0,1]^{l \times l}$, $\mathbb{P} = (p_{ij})_{i,j \in \{1,\ldots,l\}}$, and $\alpha \in [0,1]^l$ with $P(X_0 = i_0, \ldots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}$ for all $n \in \mathbb{N}_0$ and $i_0, \ldots, i_n \in E$.

Problem 4 (6 points)
Model the following system as a Markov Chain. Consider a plant species that can have either red, white or pink flowers. When crossing a red-flowered plant with a pink-flowered one, the probability that the offspring has red flowers is 1/2 and that it has pink flowers is also 1/2. White-flowered offspring is not possible. When crossing white with pink, the offspring will have white flowers with probabbility 1/2 and pink flowers with probability 1/2 while red-flowered offspring does not occur. Lastly, when crossing pink with pink, the result will be a pink-flowered plant with probability 1/2 while for both red and white flowers the probability is 1/4. You start a breeding experiment by randomly picking from a choice of a pink and a white-flowered plant and crossing it with a pink one.
Hint: Note that crossing is always done between one pink-flowered plant and one of the red-, white- or pink-flowered plants. Thus, when modeling the system one only has to focus on the line of plants where variation is possible.

A Give the Markov chain using a transition matrix and initial distribution and draw the transition graph for the Markov chain. (2 points)

B For each possible state, calculate the probability to be in that state after two steps.
If you start your experiment with 2000 pink- and 2000 white-flowered plants that you cross with pink-flowered plants and sow one resulting seed from each of the 4000 plants, how many plants of each of the flower colors would you expect to grow from the seeds? (2 points)

C Give the Markov chain in the canonical (recursive) description. (2 points)

Please hand in your solution to the problem set in the lecture on Tuesday, **October 22, 2013**. You cannot get credit for solutions turned in late and there are no exceptions from this rule. Feel free to work in groups of two and hand in a single joint solution with both names clearly written on top. Feel free to discuss with other groups, but each group has to hand in their own solution in their own words (and please do not plagiarize or copy from the internet).