Exercises for SPAß:
Statistics, Probability and Algorithms in Bioinformatics

Freie Universität Berlin, WS 2003/04
Dipl.-Math. Sven Rahmann

Problem set 13 · Handed out on 28.1.2004
Please hand in solutions by 4.2.2004 · Late solutions will not be accepted

Problem 45 (Most probable observation from a Multinomial distribution 1). Let $X = (X_1, \ldots, X_k)$ have a multinomial distribution on k categories with parameters n and $p = (p_1, \ldots, p_k) \geq 0$ with $\sum_k p_k = 1$; so $X \geq 0$, $X \in \mathbb{Z}^k$, and $\sum_k X_k = n$. Ignoring for the moment that X_k needs to be an integer, what would be the most probable observation X? Well, obviously the elements of X *are* integers. So in the next problems, we would like to know what is the most probable observation X *with* this restriction.

Problem 46 (Most probable observation from a Multinomial distribution 2). Describe an MCMC-algorithm to sample X from a specified Multinomial distribution. By introducing a temperature parameter, describe how to turn it into a simulated annealing algorithm to find the most probable observation.

Problem 47 (Most probable observation from a Multinomial distribution 3). Implement the simulated annealing algorithm to find the most probable observation X for the multinomial distribution on 10 categories with

$$p = (0.0904, 0.1164, 0.1354, 0.1085, 0.0259, 0.0596, 0.1375, 0.1347, 0.0603, 0.1313)$$

and $n = 1, \ldots, 50$ observations. Experiment with the cooling rate to get presentable results.

Problem 48 (Most probable observation from a Multinomial distribution 4). Find a deterministic algorithm that is linear in the number of categories k and polynomial in n to compute the most probable observation X.

The last problem is for extra credit. Do try it! And please tell me if you find the solution in the literature (you still get full credit — it's not that easy to find).