Quantitative Bioinformatics **Homework 4** **Due on Tuesday 03/11/2014 at 6pm**

Important Reminder: the Midterm Exam (covering Chapters 1, 2, 3, 4, 6, 7 plus all outside materials discussed) will be held on Tuesday March 18th 6.00-7.15pm. Students may bring two (2) double-sided pages of their own handwritten notes; don't forget your calculator.

Directions: Clearly, concisely and accurately answer the following exercises, writing ***very*** neatly; email submissions are not accepted – turn in a hard copy of your solutions. Keep at least 4 decimal places in your calculations.

➔ **Read Chapters 6 and 7 and also the "Ends-Space Free Algorithm" notes on our webpage** ←

1. [All Students] Suppose that we have 10 red balls, 20 blue balls and 30 yellow balls in the container. Balls are well mixed. Now draw 8 balls **without replacement**. Let $\boldsymbol{A}$ be the event that no red balls are drawn, and $\boldsymbol{B}$ the event that $\boldsymbol{k}$ blue balls are drawn (where $\boldsymbol{0 \leq k \leq 8}$). Find (a) $\boldsymbol{P(A)}$, and (b) $\boldsymbol{P(B|A)}$.

2. [All Students] Consider an occasionally dishonest casino dealer who uses two types of dice. Of the total number of dice that he uses, **99%** are fair, and the remaining **1%** are loaded in such a way that a **six** (6) comes up **50%** of the time. Suppose that we randomly choose a die from the casino table without knowing if it is fair or not. Find the following probabilities: (a) **P(six | D_{LOADED})**, (b) **P(six | D_{FAIR})**, (c) **P(six, D_{LOADED})**, (d) P**(six, D_{FAIR})**, (e) the probability of rolling a six from the die we randomly picked up, and (f) given that we roll the selected die and get a six, the probability that the die is loaded. Here, **D_{LOADED}** means that the die is loaded, etc.

3. [All Students] Using the scoring: match gives +1, mismatch gives -1, and indel gives -2, we wish to align the sequences **`CATT`** and **`GAATCT`**. Using the blank sheet given out in class on 2/25, (a) find all global alignment(s), and find all local alignment(s) of these sequences. Clearly show your final answer in both cases. Also, show all work and clearly mark and turn in your worksheets (including the relevant 'trace-back paths').

4. [All Students] Redo the alignment from the previous exercise using the same scoring and the ends-space free algorithm (ESFA); this alignment method is not covered in our text but there is a brief discussion and example on our webpage. The ESFA method is analogous to the local method in starting with zeroes in the first row and first column, but as opposed to the local method, negative values can and do occur in the center of the table. Also, with the ESFA, the "start" is always either on the left or top of the matrix, and the "end" is always on the bottom or right of the matrix. The method ends with the largest value on the bottom or right of the matrix. Clearly give all your answers and indicate each on the blank sheet provided.

5. [Graduate Students only] When sampling from the infinite population of mice, in any randomly chosen group of $\boldsymbol{N}$ mice, mice are either mutant or not. Also, regardless of a mouse's gender, the probability that the mouse is a mutant is $\boldsymbol{p}$; note that the number of mutant mice (of the $\boldsymbol{N}$ sampled) is a random variable. Further, in the randomly selected group of $\boldsymbol{N}$ mice, let $\boldsymbol{n}$ denote the number of male mice and $\boldsymbol{N - n}$ denote the number of female mice. The symbol $\boldsymbol{A}$ denotes the event that $\boldsymbol{y}$ male mice are mutants and $\boldsymbol{B}$ denotes the event that, in all, $\boldsymbol{m}$ mice are mutants. Using the definition of conditional probability, find the probability $\boldsymbol{P(A \mid B)}$. Show all work/derivations.
   Important Hint: $\boldsymbol{A \cap B}$ is the event that of $\boldsymbol{n}$ males there are $\boldsymbol{y}$ mutants and of $\boldsymbol{N - n}$ females there are $\boldsymbol{m - y}$ mutants. Note also that observing a mutant male mouse is independent of observing a mutant female mouse.

6. [Graduate Students only] How many sixes in a row would we need to see in Problem 2 above before it is more likely that we had picked a loaded die? That is, find the smallest value of $\boldsymbol{n}$ such that **P(D_{LOADED} | n sixes) exceeds ½**. Show all work.