ENGINEERING TRIPOS PART IIB

Wednesday 10 May 2006 2.30 to 4

Module 4M8

BIOINFORMATICS

*Answer not more than **two** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

There are no attachments.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
Engineering Data Book
CUED approved calculator allowed

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

1 (a) What are the differences between the PatternHunter, BLAST, Smith-Waterman and Needleman-Wunsch algorithms? [20%]

(b) Discuss the use of affine gap penalties in contrast to constant gap penalties. [20%]

(c) Describe the UPGMA (Unweighted Pair Group Method with Arithmetic mean) algorithm. [20%]

(d) What does the ultrametric property of a tree tell us about the evolutionary process? [20%]

(e) Discuss the properties and assumptions of the Jukes-Cantor and the Kimura two parameters models of DNA evolution. [20%]

2 (a) Describe the mathematical steps involved in using microarrays to find genes that are differentially expressed between two conditions. What potential problems must be overcome when looking for differentially expressed genes? [40%]

(b) What is bootstrapping, and why is it useful in microarray analysis? Give two examples of bootstrapping techniques applied to microarray analysis. [30%]

(c) Describe the linear discriminant analysis and K nearest-neighbours (Knn) classifiers. How do they compare with each other, and which would you use when classifying microarray data? [30%]

3 Consider a system with stochastic reaction events $x \xrightarrow{\lambda} x+1$ and $x \xrightarrow{\beta x^2} x-2$.

(a) Write down the master equation for the probability $p(x)$. [20%]

(b) Describe the Gillespie algorithm for generating sample paths for the system. [20%]

(c) Write down the exact differential equation for the average, $\langle x \rangle$. Approximate the equation, expressing $\frac{d\langle x \rangle}{dt}$ in terms of $\langle x \rangle$, assuming that fluctuations are negligible. [10%]

(d) Write down the stochastic reaction events of another system which has the same deterministic linearized dynamics found in part (c). Make this system differ from the original in at least two respects. [20%]

(e) Use the fluctuation dissipation theorem to estimate $\eta = \frac{\sigma_x^2}{\langle x \rangle^2}$, where σ_x^2 is the variance of x, for both the original system and the one you described for part (d). Comment on the result. [30%]

END OF PAPER