ENGINEERING TRIPOS   PART IIB

Monday 3 May 2010   2.30 to 4

Module 4G1

SYSTEMS BIOLOGY

*Answer **all** three 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

jmg02

1 (a) Discuss the Gibbs sampling algorithm in the context of identifying DNA motifs. [35%]

(b) Discuss, with one example, the information you can extract from a sequence logo. [35%]

(c) Discuss the predictability of bacterial genome motifs using Markov chain models. [30%]

2 A scientist is interested in the changes on gene expression in prostate tumours in older males introduced by the consumption of coffee and red wine. A clinical trial was performed and RNA samples were extracted from biopsies of prostate tumours of: patients who had drunk no coffee nor red wine for a month (Control), patients who had drunk two *espressos* a day but no red wine for a month (Coffee), patients who had drunk two glasses of red wine a day but no coffee for a month (Wine), and patients who had drunk both two *espressos* and two glasses of red wine a day for a month (Coffee and Wine). The samples were hybridised against a total of six two-colour microarrays. The following table summarises the experimental design, specifying the sample types hybridised against each array and the dyes used in labelling each sample:

Array number	**Cy3 (green)**	**Cy5 (red)**
1	Control	Wine
2	Wine	Coffee and Wine
3	Coffee and Wine	Coffee
4	Coffee	Control
5	Control	Coffee and Wine
6	Coffee	Wine

(a) Draw a diagram of the layout of this experimental design, using conventional arrows to represent the two-colour arrays. [20%]

(b) Determine the design matrix, taking the coffee effect, the wine effect, and the interaction between wine and coffee as your independent parameters. [25%]

(c) Determine which effect(s) this design would allow you to estimate with higher precision. Explain the caveats of these precision estimates by discussing the concept of effective replication. [35%]

(d) This design is not completely dye-balanced. Please explain why. How would you balance the design for dye if you could add two arrays and had no restrictions on the amount of any of the samples? [20%]

3 (a) Consider a birth-death process described by $x \xrightarrow{g(x)} x+1, \quad x \xrightarrow{\lambda x} x-1$.

(i) Write down the master equation for the probability of x taking value k at time t. [10%]

(ii) Derive using the master equation a differential equation for the mean of x (show detailed steps in your answer). Hence find the value of the ratio $\frac{\langle g(x)\rangle}{\langle x\rangle}$ at equilibrium. [20%]

(iii) A differential equation for $\langle x^2\rangle$ is

$$\frac{d\langle x^2\rangle}{dt} = 2\langle xg(x)\rangle - 2\lambda\langle x^2\rangle + 2\lambda\langle x\rangle$$

Using the approximation

$$g(x) \approx g(\langle x\rangle) + g'(\langle x\rangle)(x - \langle x\rangle)$$

derive a differential equation for the variance of x.
Note: $g'(\langle x\rangle)$ denotes the derivative $\frac{dg(x)}{dx}$ evaluated at $x = \langle x\rangle$. [30%]

(b) (i) Consider a gene which is turned on and off with rates

$$\mathrm{ON} \underset{\beta_2}{\overset{\beta_1 x}{\rightleftarrows}} \mathrm{OFF}$$

where x is the number of molecules of a species that is affecting the expression of the gene. Show that at equilibrium the probability of the ON state is $P(\mathrm{ON}) = \frac{\beta_2}{\beta_2+\beta_1 x}$. [10%]

(ii) Consider now the birth-death process in (a) with x denoting the number of mRNAs of a gene that is expressed. The mRNA is inhibiting its own transcription with $g(x)$ given by the expression for $P(\mathrm{ON})$ in (b)(i).
Using the approximation in (a)(iii), derive an expression for the variance of x at equilibrium. [20%]

(iii) Discuss when the approximations made in (b)(ii) are valid. [10%]

END OF PAPER