Introduction to bioinformatics, Autumn 2007, Exercise 2

15.9.2007

1. (Chapter 2, Exercise 4) Suppose N has a binomial distribution with $n = 10$ and $p = 0.3$.

   (a) Using the formula (2.17), calculate $P(N = 0)$, $P(N = 2)$, $E(N)$ and $VarN$.

   (b) Using R and Computational Example 2.2., simulate observations from N. Use the simulated values to estimate the probabilities you calculated in (a), and compare with the results in (a).

   (c) Now use R to simulate observations from N when $n = 1000$ and $p = 0.25$. What is your estimate of $P(N \geq 280)$? (See (2.20).)

2. (Chapter 2, Exercise 5) Verify the terms in the first row of the transition matrix P presented in Section 2.6.3. Describe how you would use the sequence of *M. genitalium* to produce this matrix.

P =

	A	C	G	T
A	0.423	0.151	0.168	0.258
C	0.399	0.184	0.063	0.354
G	0.314	0.189	0.176	0.321
T	0.258	0.138	0.187	0.415

3. (Chapter 2, Exercise 6) Find the stationary distribution of the chain with transition matrix P in Section 2.6.3; that is, solve the equations $\pi = \pi P$ subject to the elements of π begin positive and summing to 1. Compare π to the base composition of *M.genitalium*, and comment.

4. (Chapter 2, Exercise 12 (a) – (c)) In this exercise we have two random variables X and Y which are not independent. Their joint probability distribution is given in the following table:

		Y 1	3	6	9
	2	0.11	0.05	0.20	0.08
X	3	0.20	0.02	0.00	0.10
	7	0.00	0.05	0.10	0.09

   The values of X are written in the first column and the values of Y in the first row. The table is read as $P(X = 7 \& Y = 6) = 0.10$, and so on.

(a) Find the marginal distribution of X and Y. (That is, $P(X = 2), P(X = 3), \ldots$)

(b) Write $Z = XY$. Find the probability distribution of Z.

(c) The **covariance** between any two random variables is defined by

$$Cov(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Show that $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}X \times \mathbb{E}Y$.

5. (Chapter 2, Exercise 12 (d) – (f))

(a) Find $\mathbb{E}X, \mathbb{E}Y, \sigma_X^2 = VarX, \sigma_Y^2 = VarY$, and $Cov(X,Y)$ for the example in the table.

(b) The **correlation coefficient** ρ is defined by $\rho_{X,Y} = Cov(X,Y)/\sigma_X\sigma_Y$. It can be shown that $-1 \leq \rho \leq 1$, the values ± 1 arising when Y is a linear function of X. Verify this last statement.

(c) Calculate ρ for the example in the table.