Some exam-like questions.

All questions in the exam will contain a biological component (e.g. explain the central dogma of molecular biology) which will require a brief essay-like answer and will account for about a third of the marks assigned to the question. In this sheet we focus on revising the informatics components of the course so I will skip the biological parts of the questions. You are (and will be in the exam) allowed to use a calculator (but will not be allowed the lecture slides of course).

1. *a*) Define the concept of expectation, and the finite sample estimates of mean, covariance and correlation between random variables. Introduce Pearson's correlation between two vector measurements $\mathbf{x}_1$ and $\mathbf{x}_2$ and show that it is indeed obtained as the correlation between the two random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ if they are assumed to have zero mean.

   b) You are given the following five profiles from a microarray experiment:

$$\begin{aligned}
\mathbf{x}_1 &= [1.1, 3.2, 3.5, 2.4, 0.7] \\
\mathbf{x}_2 &= [0.6, 1.5, 1.9, 1.4, 0.3] \\
\mathbf{x}_3 &= [1.5, 2.9, 3.7, 2.2, 0.8] \\
\mathbf{x}_4 &= [1.7, 0.2, 0.5, 1.4, 1.7] \\
\mathbf{x}_5 &= [1.2, 0.2, 0.3, 0.8, 1.1].
\end{aligned}$$

   Using Pearson's correlation as a similarity measure, hierarchically cluster these profiles.

2. *a*. Define the concept of network motif, and explain the Erdos Renyi construction of random graphs.

   b The joint distribution of four Gaussian random variables is given by

$$p(x_1, \ldots, x_4) \propto \exp\left[-\frac{1}{2}\left(c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + c_{44}x_4^2 + c_{13}x_1x_3 + c_{14}x_1x_4 + c_{23}x_2x_3 + c_{24}x_2x_4\right)\right].$$

   Determine the conditional distribution $p(x_1|x_2, x_3, x_4)$. Draw a network representing the dependencies between these random variables.

3. *a*. Let x be distributed according to a Gaussian with mean μ and variance σ^2, and let μ be given a Gaussian prior distribution with mean m and variance s^2. Given observations $x_1, \ldots, x_N$, compute the posterior distribution over μ.

   b The general formula for the computation of the forward message in a LDS is given by

$$\alpha(x(t)) = \int dx(t-1)\alpha(x(t-1))\mathcal{N}(x(t)|Ax(t-1), \Sigma_w)\,\mathcal{N}(y(t)|Bx(t), \Sigma_\varepsilon).$$

   Let x and y be 1 dimensional; $\alpha(x(t)) = \mathcal{N}(\mu(t), \sigma_t^2)$. Compute the expression for the mean and variance of the forward message at time t in terms of the mean and variance at time $t-1$ and of the parameters of the model $A, B, \Sigma_w, \Sigma_\varepsilon$.