Introduction to bioinformatics, Autumn 2006, Exercise 2

29.9.2006

1. (Chapter 3, Exercise 7) Suppose that Z has a standard normal distribution.
   (a) Find $P(-1 \leq Z \leq 1), P(-1 \leq Z \leq 2), P(-2 \leq Z \leq -1), P(-\infty \leq Z \leq 1)$.
   (b) If $P(Z \leq a) = 0.45$ and $P(0 \leq Z \leq b) = 0.45$, find a and b. [Hint: Use qnorm in R.]
2. (Chapter 3, Exercise 8) In a certain genome the bases appear to be iid and $p_g = 0.3$. Define the (binomial) count of the number of Gs in the first 1000 bases as $N = X_1 + X_2 + \cdots + X_{1000}$.
   (a) Give the mean and variance of N.
   (b) Approximate, using the Central Limit Theorem, $P(0 \leq N \leq 329)$ and $P(285.5 \leq N \leq 329)$.
   (c) Produce a histogram for 1000 replicates of N and compare the results with those of (b).
3. (Chapter 3, Exercise 10) Suppose N is binomial with $n = 1000$ and success probability p. Find a 90% confidence interval for p using $\hat{p} = N/1000$ when $N = 330$. What are $\mathbb{E}\hat{p}$ and Var $\hat{p}$?
4. Perform local alignment of the sequences

   ```
   s = TCAGTTGCCAAACCCGCT
   t = AGGGTTGACATCCGTTTT
   ```

   with $\mu = 1$, $\delta = 2$ and uniform match score 1 by constructing the dynamic programming matrix.
5. Perform global alignment of the sequences

   ```
   s = TCAGTTGCTCA
   t = AGGGTTGACA
   ```

   with $\mu = 1$, $\delta = 2$ and uniform match score 1 by constructing the dynamic programming matrix.