

**MATH 3900, Spring 2009**  
**Homework 12**  
**Due 04/24/09**

1. HIV RNA consists of  $10^4$  bases. Probability of an error during replication is  $3 * 10^{-5}$  per base per replication cycle. From class: there are about  $10^8$  of new fully functional viral particles produced per day.

- a) What is the total expected number of errors per replication cycle?
- b) Find probability to get a mutant with mutation in exactly 1 base (use appropriate approximation as needed). What number of these mutants do you expect to find among the population of viruses produced on a given day?
- c) Number of all possible 1-base mutants is  $3 * 10^4$ . Let's say that one of these mutations leads to antibiotic resistance. Assuming that all viruses produced on a given day are different (or as diverse as they can be), what is the probability that antibiotic resistance will appear on a given day? d) Repeat b-c with two-base-mutations. Number of all possible two base mutants is approximately  $4.5 * 10^8$ .

When you are done, notice that even though the probability that you find to have a given two-base mutation in an infected person in one day is relatively small, it will only take a population of 136 people to have the expected number of antibiotic resistance mutations equal to 1. So, even if you need two mutations for antibiotic resistance (or resistance to therapy with 2 drugs), it will show up quickly.

- e) Now repeat b-c with three-base-mutations. These will be very rare, so the resistance that needs three mutations is hard to come by, hence the combination drug therapy.

2. **(computing)**  $X$  is Normally distributed  $N(2,4)$ . Find  $\Pr(X < 0.5)$ ,  $\Pr(X > 3)$ ,  $\Pr(0 < X < 4)$ . Use Matlab function `normcdf(x)` to give you probabilities  $\Pr(Z < x)$ .

3. **(computing)** a) Produce 1000 exponentially distributed numbers with  $\lambda = 1$ . Put them into 20 bins. Plot the histogram. (you can use matlab function `hist(T,M)`, where  $T$  is the times tha you generated and  $M$  is the number of bins. To look at its help file type 'help hist', to plot the result use `bar(hist(T,M))`)  
b) Produce 1000 pairs of exponentially distributed numbers with  $\lambda = 1$ , and take the average of each pair. Put these values into bins and plot the histogram.  
c) Produce 1000 sets of 10 exponentially distributed numbers with  $\lambda = 1$ , and take the average of each set. Put these values into bins and plot the histogram.  
d) Produce 1000 sets of 30 exponentially distributed numbers with  $\lambda = 1$ , and take the average of each set. Put these values into bins and plot the histogram. Do these results look more and more like normal distribution? What are the  $\mu$  and  $\sigma^2$ ? Mark the mean and the mean plus and minus sigma on the normal distribution.

4. Us ethe normal approximation to the binomial with and without continuity correction to find he following probabilities. Noice how much difference

the correction makes.

- a) Ten percent of people are known to carry a certain gene. In a sample of 200, what is the probability that between 5 and 15 carry the gene? (to compare: the exact answer from the binomial is 0.1431)
- b) Thirty percent of cells in a small organism are not functioning. What is the probability that an organ consisting of 250 cells is functioning if it requires 170 cells to work? (to compare: the exact answer from the binomial is 0.777)