Homework #2

1. (6 points) There are about 1000 ribosomes in E. coli cell. Each ribosome generates proteins at a rate of 12 amino acids/sec. An average protein has 300 amino acids.
http://bionumbers.hms.harvard.edu/bionumber.aspx?&id=107785&ver=3&trm=ribosome%20translation%20rate

(a) Calculate the rate λ at which all ribosomes in the cell combined crank up the proteins

$$\text{lambda}=1000*12/300=40$$

Answer: 40 proteins/sec

(b) If one observes the cell during 0.1 seconds what is the probability that at least one new protein will be produced?

$$1-\exp(-rx)=1-\exp(-4)=0.9817$$

Answer: 0.9817

(c) That exactly 4 proteins will be produced?

$$(xr)^4/4!*\exp(-xr)=0.1954$$

Answer: 0.1954

2. (6 points) If the average number of claims handled daily by an insurance company is 5 and the distribution of the daily number of claims is Poisson, what is the probability that there will be 4 claims every day in exactly 3 of the next 5 days? Assume that the number of claims on different days is independent.

Answer: 0.0367

$$\lambda=5$$
$$P(x=4)=(5^4)/4!*\exp(-5)$$
$$P(\text{event})=C\_5^3*P(x=4)^3*(1-P(x=4))^2$$

3. (12 points) Sequencing technologies can only "read" short fragments from a genome. Given that the process through which the sequences are generated is random, it is possible that certain parts of the genome will remain uncovered unless an impractical amount of sequences are generated.
We know that the size of the human genome is 3×10^9 bp. Now a new human genome has been sequenced and it's randomly covered by 30×10^6 reads (read length is 300 bp). We assume that the number of times a base in the human genome is covered follows a Poisson distribution.

(a) What is the probability that a particular base is not covered by any read?

$$\lambda=np=30*10^6*300/(3*10^9)=3$$
$$P(x=0)=\exp(-3)$$

Answer: 0.0498

(b) One randomly picks bases in this genome one at a time. What is the expected number of bases one has to pick at before the first uncovered base is identified

$$\text{geometric dist: } E(X)=1/p=1/0.0498$$

(c) What is the expected number of bases one has to look at before ten uncovered bases are identified?

negative binomial dist: $E(X)=r/p=10/0.0498$

4. (10 points) Assume X is normally distributed with a mean of 5 and a standard deviation of 4.

(a) Determine $P(X > 0)$

$P(X>0)=P(Z=x-5/4>-1.25)=P(Z<1.25)=$
0.8943

(b) Determine $P(2 < X < 9)$

$P(-0.75<Z<1)=P(Z<1)-P(Z<-0.75)=P(Z<1)-1+P(Z<0.75)=0.6148$

(c) If $P(x < X < 9) = 0.2$, what is x?

$P(x-5/4<Z<1)=0.2=P(Z<1)-P(Z<x-5/4)$
$p(Z<x-5/4)=0.6413 => Z<0.36 =>x=6.44$

5. (8 points) The annual rainfall (in inches) in a certain region is normally distributed with mean $\mu = 30$, and standard deviation $\sigma = 4$. What is the probability that in 2 of the next 5 years the rainfall will exceed 34 inches? (Assume that the rainfalls in different years are independent.)

$P(X>34)=1-P(X<34)=1-P(Z<1)=0.158$
$5!/2!3!(0.158)^2(1-0.158)^3=0.15$

6. (8 points) Measurement error that is normally distributed with a mean of zero and a standard deviation of 0.5 grams is added to the true weight of a sample. Then the measurement is rounded to the nearest gram. Suppose that the true weight of a sample is 165.5 grams.

    (a) What is the probability that the rounded result is exactly 167 grams?

$$P(166.5 < X < 167.5) = P(2 < Z < 4) = P(Z < 4) - P(Z < 2) = 0.9999 - 0.9772 = 0.227$$

Answer: 0.0227

    (b) What is the probability that the rounded result is 167 grams or greater?

$$P(X > 166.5) = P(Z > 2) = 1 - P(Z < 2) = 0.228$$

Answer: 0.0228