Homework #2

Please present 4 significant figures in your final answers for probabilities

1. (6 points) There are about 1000 ribosomes in E. coli cell. Each ribosome generates proteins at a rate of 12 amino acids/sec. We assume a that protein production is a constant rate (Poisson) process. 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

Answer: $12 \times \frac{1000}{300}$ =40 proteins/sec

(b) If one observes the cell for 0.1 seconds what is the probability that at least one new protein will be produced?
Answer: On average, there are 4 proteins that are produced during the time window of 0.1 sec.
$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4} = 0.9817$

(c) What is the probability that exactly 7 proteins will be produced?
Answer: $P(X = 7) = \frac{4^7 e^{-4}}{7!} = 0.05954$

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 at least 2 of the next 5 days? Assume that the number of claims on different days is independent.

Answer: The probability that there will be 4 claims on any day is P(X=4) = $\frac{5^4 e^{-5}}{4!} = 0.17547$.
Therefore, the probability that there will be 4 claims every day in at least 2 of the next 5 days is
1 – P(Y=0) – P(Y=1) = 1 - $C_0^5 0.17547^0 (1 - 0.17547)^5 - C_1^5 0.17547^1 (1 - 0.17547)^4$ = 1 - 0.38109 - 0.40551 = 0.2134

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?

Answer: the average time a base is covered is $30 \times 10^6 \times \frac{300}{3\times10^9} = 3$. The probability that a particular base is not covered by any read is $P(X=0) = e^{-3} = 0.04979$

(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?

Answer: $\frac{1}{0.0498} = 20.08 \approx 20$

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

Answer: $\frac{10}{0.0498} = 200.80 \approx 201$

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

(a) Determine P(X > 2)

Answer: $P(X > 2) = 1 - P(X \le 2) = 1 - 0.22663 = 0.7734$

(b) Determine P(0 < X < 9)

Answer: $P(0 < X < 9) = P(X < 9) - P(X < 0) = 0.84134 - 0.10565 = 0.7357$

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

Answer: If $P(x < X < 9) = 0.2$, then $P(X < x) = 0.6413$. By looking at the table, we know that $x = 6.448$.

5. (8 points) The annual rainfall (in inches) in a certain region is normally distributed with mean μ= 30, and standard deviation σ = 4. What is the probability that in 3 of the next 5 years the rainfall will exceed 34 inches? (Assume that the rainfalls in different years are independent.)

Answer: The probability that the rainfall exceed 34 inches is $P(X > 34) = 1 - P(X \leq 34) = 0.1587$. Therefore, the probability that in 3 of the next 5 years the rainfall will exceed 34 inches is $P = C_3^5 0.1587^3(1 - 0.1587)^2 = 0.02829$

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 156.5 grams.

(a) What is the probability that the rounded result is exactly 158 grams?

Answer: Let us denote X as measurement and ϵ as error. Then, we have $X = 156.5 + \epsilon$. The probability that the rounded result is exactly 158 grams is $P(157.5 \leq X < 158.5) = P(X < 158.5) - P(X < 157.5) = P(\epsilon < 2) - P(\epsilon < 1) = 0.0227$

(b) What is the probability that the rounded result is 158 grams or greater?

Answer: $P(X \geq 157.5) = P(\epsilon \geq 1) = 1 - P(\epsilon < 1) = 0.0228$