

Midterm Exam

Name: _____

1. **(15 points)** If the letters of ILLINI are randomly ordered, all orderings being equally likely, what is the probability that not a single position has the same letter as in the original order? Hint: 3 Is (and 2 Ls) are identical.

2. **(10 points)** A novel biomarker diagnosis assay is 95% effective in detecting a certain disease when it is present. The test also yields 2% false-positive result. If 1% of the population has the disease, what is the probability that a randomly chosen member of the population has the disease given that his/her assay result is positive?

3. **(15 points)** Sequencing technologies can “read” many short fragments (simply called reads) from a genome. Given that the process through which the read 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. Human genome is 3×10^9 bp long. A patient’s genome has been sequenced and it is randomly covered by 300×10^5 reads (each read is 100 bp long). We assume that the number of times a base in the human genome is covered follows a Poisson distribution.
 - (a) (5 points)** What is the probability that a particular base is not covered by any reads?

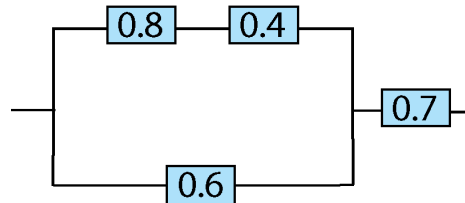
 - (b) (5 points)** What is the probability that a particular base is covered by at least two reads?

 - (c) (5 points)** We now start randomly selecting bases in our genome. What is the expected number of bases we have to look at before exactly 4 such uncovered bases are identified?

4. **(10 points)** The common logarithm (base 10) of the expression level (mRNA copies/cell) of a cancer driver gene in a randomly selected cell is normally distributed with mean $\mu = 4$, and standard dev. $\sigma = 0.5$.
 - (a) (5 points)** What is the probability that the expression level measured in a given cell is between 1000 and 100,000 mRNA copies/cell?

(b) (5 points) Expression level of this gene was measured in 6 individual cells. What is the probability that in exactly 3 cells gene's expression level within these bounds?

5. **(10 points)** The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



6. **(10 points)** In a data communication system, several messages that arrive at a node are bundled into a packet before they are transmitted over the network. Assume the messages arrive according to a Poisson process with the mean rate equal to one message per two minutes. Five messages are required to form a packet and the packet is formed immediately after the last message has arrived.
- (a) (5 points)** What is the probability that a time interval between two consecutive messages is longer than 4 minutes?

(b) (5 points) What is the mean time until a packet is formed, that is, until exactly five messages have arrived at the node?

7. **(10 points)** The joint probability mass function of discrete random variables X and Y taking values $x = 1, 2$ and $y = 1, 2$ respectively, is given by $f_{XY}(x,y) = c \cdot x \cdot y$. Find
- (a) the normalization constant c
 - (b) the marginal probability of $Y=1$
 - (c) the conditional probability of $Y=2$ given that $X = 1$
 - (d) $\text{Cov}(X,Y)$
 - (e) are X and Y independent? Please justify.