Midterm will be held in this room this Thursday 3/14 during regular class hours 9:30am-10:50am
Midterm Info

- No office hours since no one selected to come
- Closed book exam; no books, notes, laptops, phones...
- Calculators (not on smartphones) can be used.
- You can prepare one 2-sided cheat sheet
- The following two printouts will be provided
<table>
<thead>
<tr>
<th>Name</th>
<th>Probability Distribution</th>
<th>Mean</th>
<th>Variance</th>
<th>Section in Book</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrete</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>( \frac{1}{n}, \ a \leq b )</td>
<td>(( b + a ))/2</td>
<td>(( b - a + 1 ))^2 - 1</td>
<td>3-5</td>
</tr>
<tr>
<td>Binomial</td>
<td>( \binom{n}{x} p^x(1 - p)^{n-x} ), ( x = 0, 1, \ldots, n ), ( 0 \leq p \leq 1 )</td>
<td>np</td>
<td>np(1 - p)</td>
<td>3-6</td>
</tr>
<tr>
<td>Geometric</td>
<td>( (1 - p)^{x-1} p ), ( x = 1, 2, \ldots, 0 \leq p \leq 1 )</td>
<td>1/p</td>
<td>(1 - p)/p^2</td>
<td>3-7.1</td>
</tr>
<tr>
<td>Negative binomial</td>
<td>( \binom{x - 1}{r - 1} (1 - p)^{r-1} p^r ), ( x = r, r + 1, r + 2, \ldots, 0 \leq p \leq 1 )</td>
<td>( r/p )</td>
<td>( r(1 - p)/p^2 )</td>
<td>3-7.2</td>
</tr>
<tr>
<td>Hypergeometric</td>
<td>( \binom{k}{x} \binom{n-k}{n-x} ) ( \binom{n}{x} ), ( x = \max(0, n - N + K), 1, \ldots, \min(K, n), K \leq N, n \leq N )</td>
<td>np, where ( p = \frac{K}{N} )</td>
<td>np(1 - p) ( \frac{N - n}{N - 1} )</td>
<td>3-8</td>
</tr>
<tr>
<td>Poisson</td>
<td>( \frac{e^{-\lambda} \lambda^x}{x!} ), ( x = 0, 1, 2, \ldots, 0 &lt; \lambda )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
<td>3-9</td>
</tr>
<tr>
<td><strong>Continuous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>( \frac{1}{b - a}, \ a \leq x \leq b )</td>
<td>(( b + a ))/2</td>
<td>(( b - a ))^2</td>
<td>4-5</td>
</tr>
<tr>
<td>Normal</td>
<td>( \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} )</td>
<td>( \mu )</td>
<td>( \sigma^2 )</td>
<td>4-6</td>
</tr>
<tr>
<td>Exponential</td>
<td>( \lambda e^{-\lambda x}, 0 \leq x, 0 &lt; \lambda ), ( x ) ( \lambda^2 )</td>
<td>( 1/\lambda )</td>
<td>( 1/\lambda^2 )</td>
<td>4-8</td>
</tr>
<tr>
<td>Erlang</td>
<td>( \frac{\lambda x^{r-1} e^{-\lambda x}}{(r - 1)!}, 0 &lt; x, r = 1, 2, \ldots )</td>
<td>( r/\lambda )</td>
<td>( r/\lambda^2 )</td>
<td>4-9.1</td>
</tr>
<tr>
<td>Gamma</td>
<td>( \frac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 &lt; x, 0 &lt; r, 0 &lt; \lambda )</td>
<td>( r/\lambda )</td>
<td>( r/\lambda^2 )</td>
<td>4-9.2</td>
</tr>
</tbody>
</table>
What is included in Midterm?

• Probability of events (set operations), Multiplication rules. Combinatorics
• Bayes Theorem
• Discrete Random Variables
• Continuous Random Variables
• Other topics covered (see HW1-HW2 for inspiration)
• No Matlab exercises (since no computers)
Probability Multiplication Rules
Combinatorics
Mr. Jones has 6 different books that he is going to put on his bookshelf. Of these, 3 are chemistry books, 2 are physics books, and 1 is a mathematics book. Jones wants to arrange his books so that two conditions are met:

(1) all the books dealing with the same subject are together on the shelf

AND

(2) all chemistry books are on the leftmost side.

How many such different arrangements are possible?
Mr. Jones has 6 different books that he is going to put on his bookshelf. Of these, 3 are chemistry books, 2 are physics books, and 1 is a mathematics book. Jones wants to arrange his books so that two conditions are met:

(1) all the books dealing with the same subject are together on the shelf

AND

(2) all chemistry books are on the leftmost side.

How many such different arrangements are possible?

Answer: \((3! \times 2! \times 1!) \times 2! = 24\)
4. (4 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?
4. (4 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?

\[ 1 - (1 - 0.8) \cdot (1 - 0.4) = 0.82 \]

\[ 0.8 \quad 0.5 \]

\[ 0.1 \]

\[ 0.3 \quad 0.9 \quad 0.7 = 0.19 \]

\[ 0.87 \times 0.5 = 0.44 \]

\[ 0.44 \quad 0.19 \]

\[ 1 - (1 - 0.44) \cdot (1 - 0.19) = 0.52 \]
4. **(4 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?

![Circuit Diagram]

**Answer:** $P(\text{Operate}) = 1 - (1 - 0.3 \times 0.9 \times 0.7) \times (1 - 0.5 \times (1 - (1 - 0.8) \times (1 - 0.1))) = 0.52$
Bayes theorem
In answering a question on a multiple-choice test, a student either knows the answer or he guesses. Let $\frac{1}{3}$ be the probability that he knows the answer. If he does not know the answer, he randomly guesses one out of 4 multiple choice questions. What is the conditional probability that a student knew the answer to a question given that he answered it correctly?

A. $\frac{1}{4}$  
B. $\frac{1}{3}$  
C. $\frac{2}{3}$  
D. $\frac{1}{5}$  
E. I don’t know

Get your i-clickers
In answering a question on a multiple-choice test, a student either knows the answer or he guesses. Let \( \frac{1}{3} \) be the probability that he knows the answer. If he does not know the answer, he randomly guesses one out of 4 multiple-choice questions. What is the conditional probability that a student knew the answer to a question given that he answered it correctly?

**Answer:** \( P(K) = \frac{1}{3}, \ P(K') = \frac{2}{3}, \ P(C|K) = 1, \ P(C|K') = \frac{1}{4}. \)

\[
P(K|C) = \frac{P(C|K) \cdot P(K)}{P(C)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3}} = \frac{2}{3} = 0.666...\]
Discrete Probability Distributions
What is $X$ in this problem?

• What is the random variable: Look for keywords:
  – Find the probability that....
  – What is the mean (or variance) of...

• What are parameters? Look for keywords:
  – Given that...
  – Assuming that...

3. Find $x$. Here it is.
Guide to probability distributions

- **Binomial**: # of samples, \( n \), is fixed, # of successes, \( x \), is variable
  \[
  P(X = x) = \frac{n!}{x!(n-x)!} \ p^x \ (1-p)^{n-x}
  \]

- **Geometric**: # of samples, \( x \) is variable. # of successes \( k \) is fixed. Success comes in the end
  \[
  P(X = x) = (1-p)^{x-1} \cdot p
  \]

- **Negative binomial**: # of samples, \( x \) is variable. # of successes, \( r \), is fixed
  8th success in the end
  \[
  P(X = x) = \frac{(x-1)!}{(r-1)!(x-r)!} \cdot p^r \ (1-p)^{x-r}
  \]
Poisson distribution in genomics

- $G$ - genome length (in bp)
- $L$ - short read average length
- $N$ – number of short read sequenced
- $\lambda$ – sequencing redundancy = $LN/G$
- $x$ - number of short reads covering a given site on the genome

\[
P(x) = \frac{\lambda^x e^{-\lambda}}{x!}
\]

Ewens, Grant, Chapter 5.1

Poisson as a limit of Binomial. For a given site on the genome for each short read $\text{Prob(site covered)}: p = L/G$ is very small.
Number of attempts (short reads): $N$ is very large. Their product (sequencing redundancy): $\lambda = NL/G$ is $O(1)$. 
Probability that a base pair in the genome is not covered by any short reads is 0.1. One randomly selects base pairs until exactly 5 uncovered base pairs are found. Which discrete probability distribution describes the number of attempts?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers
Probability that a base pair in the genome is not covered by any short reads is 0.1.

One randomly selects base pairs until exactly 5 uncovered base pairs are found.

Which discrete probability distribution describes the number of attempts?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers
Probability that a base pair in the genome is not covered by any short reads is 0.1

One randomly selects base pairs until exactly 5 uncovered base pairs are found.

What are the values of p, r?

A. p=0.5, r=5
B. p=0.1, r=0.5
C. p=0.1, r=5
D. p=0.5, r=0.1
E. I have no idea

Get your i-clickers
Probability that a base pair in the genome is not covered by any short reads is 0.1
One randomly selects base pairs until exactly 5 uncovered base pairs are found.

What are the values of p, r?

A. p=0.5, r=5
B. p=0.1, r=0.5
C. p=0.1, r=5
D. p=0.5, r=0.1
E. I have no idea

Get your i-clickers
Cancer happens when the gene p53 mutates. Probability of p53 to mutate per year is 5%. How many years before a patient gets disease? Which discrete probability distribution would you use to answer?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers
Cancer happens when the gene p53 mutates. Probability of p53 to mutate per year is 5%.

How many years before a patient gets disease?

Which discrete probability distribution would you use to answer?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers
Continuous Probability Distributions
2. (8 points) The length of stay at a specific emergency department in Phoenix, Arizona, in 2009 had a mean of 4.6 hours with a standard deviation of 2.9. Assume that the length of stay is normally distributed.

(A) (4 points) What is the probability of a length of stay greater than 10 hours?

(B) (4 points) How long does one have to stay in this emergency room to know that approximately 25% of all visits last even longer?
2. (8 points) The length of stay at a specific emergency department in Phoenix, Arizona, in 2009 had a mean of 4.6 hours with a standard deviation of 2.9. Assume that the length of stay is normally distributed.

(A) (4 points) What is the probability of a length of stay greater than 10 hours?

Answer: \( \frac{10-4.6}{2.9} = 1.86 \) Using table one finds Prob=1-0.9687=0.0313

(B) (4 points) How long does one have to stay in this emergency room to know that approximately 25% of all visits last even longer?

Answer: Using table one finds \( P(Z<0.67) = 0.75 \) meaning it is \( 4.6 + 2.9 \times 0.67 = 6.543 \)
1. (8 points) The expression level of a *TP53* tumor suppressor gene in a randomly selected cell is normally distributed with mean $\mu = 20$, and standard deviation $\sigma = 8$.
   
   (A)(4 points) What is the probability that the expression level in a given cell will be between 24 and 16?

   (B)(4 points) How many cells does one have to sample (on average) until there will be exactly 2 cells with such “close to average” *TP53* expression?
1. **(8 points)** The expression level of a TP53 tumor suppressor gene in a randomly selected cell is normally distributed with mean $\mu = 20$, and standard deviation $\sigma = 8$.

   (A) **(4 points)** What is the probability that the expression level in a given cell will be between 24 and 16?
   
   **Answer:** Using table one finds $\text{Prob}(Z<0.5)=0.6914$. Thus the answer is $0.6914-(1-0.6914)=0.3829$

   (B) **(4 points)** On average, how many cells does one have to sample until there will be exactly 2 cells with such “close to average” TP53 expression?
   
   **Answer:** Using the negative binomial distribution one gets $2./0.3829=5.22$
I can show you how to solve any HW1-HW2 problem.

Which one do you choose?