SFU

Math 468, Spring 2023

Midterm

March 1, 2024, 8:30 – 9:20

SFU Email:	@SFU.CA	Signature:	

First Name:	Solutions
Last Name:	
SFU ID #:	

1. Do not open this booklet until told to do so.
2. Write your name, SFU student number and email ID in the space provided.
3. Write your answer in the space provided. If additional space is needed use the back of the previous page. Your final answer should be simplified as far as is reasonable; you may leave answers in "calculator ready"expressions: such as $3 + \ln 7$ or $e^{\sqrt{2}}$.
4. To receive full credit for a particular question your solution must be complete and well presented.
5. **No** books, papers, or electronic devices other than your calculator and "Crib Sheet" can be used during the examination.
6. **During the examination, copying from, communicating with, or deliberately exposing written papers to the view of other examinees is forbidden.**

Question:	1	2	3	4	5	Total
Points:	4	8	3	7	6	28

Math 468, Spring 2023
Midterm

[4] 1. Suggest a probability distribution describing the following experimental trials. Use each distribution at most once and list a) whether it is a discrete or a continuous distribution and b) the domain of the distribution.

-1 per row w/ error up to -4

Experiment	Distribution	Discrete vs. Continuous	Domain of Distribution
The weight, X, of baby koalas	Normal	Cont	$[0, \infty)$
The number, X, of red "mutant" flowers in a population	Binomial	Discrete	$\mathbb{N}$
The time, X, until the next volcanic eruption	Exponential	Cont	$[0, \infty)$
The number, X, of rabbits caught by a hare in a year	Poisson	Discrete.	$\mathbb{N}$
The time, X, until newly planted seedlings reach maturity	Erlang or Gamma	Cont.	$[0, \infty)$
The outcome, X, of a paternity test	Bernoulli	Discrete	$\{0, 1\}$
The time, X, in seconds right now	Uniform	Cont	$[0, 60]$

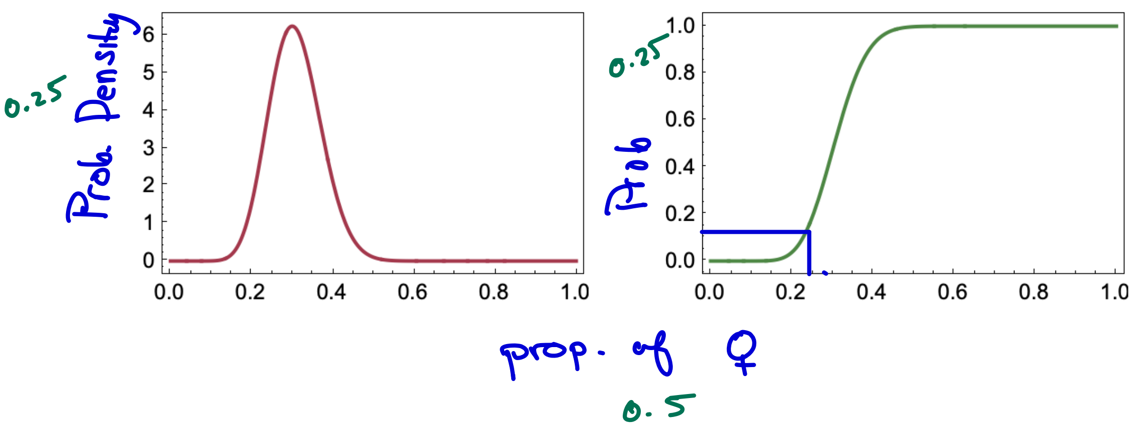
Math 468, Spring 2023
Midterm

2. In most biological systems the sex ratio (males to females) is approximately 50/50. However, in Green Sea Turtles sex is determined by egg temperature and can deviate significantly from this even ratio. A researcher surveys 50 turtle hatchlings 35 of which are male and 15 of which are female.

[1] (a) What distribution describes the true sex ratio given the observed data?

Beta Dist

[1] (b) The PDF and CDF for the distribution in part (a) are shown below. Label the axes of the plots and indicate an "even" sex ratio.



[1] (c) Using the figure above, approximate the probability that the proportion of males is greater than 75%.

0.5 0.5

Pr(♂ > 0.75) = Pr(♀ < 0.25) ≈ 0.1

Math 468, Spring 2023
Midterm

[1] (d) What is the expected sex ratio in the population as a whole given the observed data?

$\alpha = 15+1 \qquad \beta = 35+1$

$$E[x] = \frac{\alpha}{\alpha+\beta} = \frac{16}{16+35} = 0.307$$

(0.5) (0.5)

$Pr(♀) = 0.307$

$Pr(♂) = 0.693$

[2] (e) Turtles are long lived with 15% of turtles living to age 70. Suppose that 10% of male turtles live to age 70. Using the sex ratio in part (d), what is the probability that a female turtle lives to age 70?

$Pr(old) = 0.15$

$Pr(old \mid ♂) = 0.1$

(1pt) $Pr(old) = Pr(old \mid ♂) Pr(♂) + Pr(old \mid ♀) Pr(♀)$

$0.15 = 0.1 \cdot 0.693 + x \cdot 0.307$

$$Pr(old \mid ♀) = \frac{0.15 - 0.0693}{0.307} = 0.263$$

(1pt)

[2] (f) A researcher finds a 70 year old turtle, what is the probability that that turtle is male?

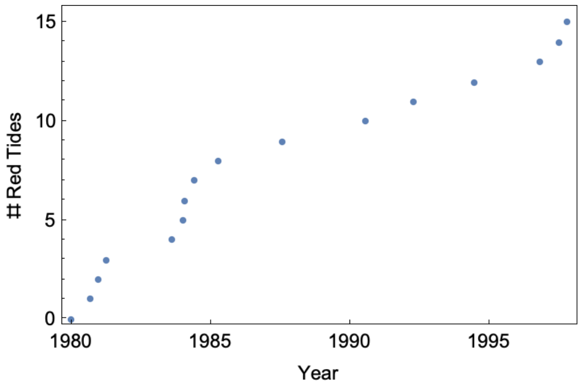
$$Pr(♂ \mid old) = \frac{Pr(old \mid ♂)\, Pr(♂)}{Pr(old)} = \frac{0.1 \cdot 0.693}{0.15}$$

(1pt) (1pt)

$= 0.462$

Math 468, Spring 2023
Midterm

3. "Red tides" are the rapid population growth of one of several different types of algae. Climate change is changing the rate at which red tides occur. An inter-tidal researcher is studying the occurrence of red tides. Their data on the occurrence of red tides from 1980 to 2000 is shown below:



[1] (a) Is the occurrence of red tides appropriately modeled as Markovian, why or why not?

Yes: Whether a red tide occurs in the future prob. depends predominantly on if there is currently a Red tide (time since last red tide) not the tides before that

0.5 explination

[1] (b) Is the occurrence of red tides appropriately modeled as time-homogeneous, why or why not?

No: Climate change

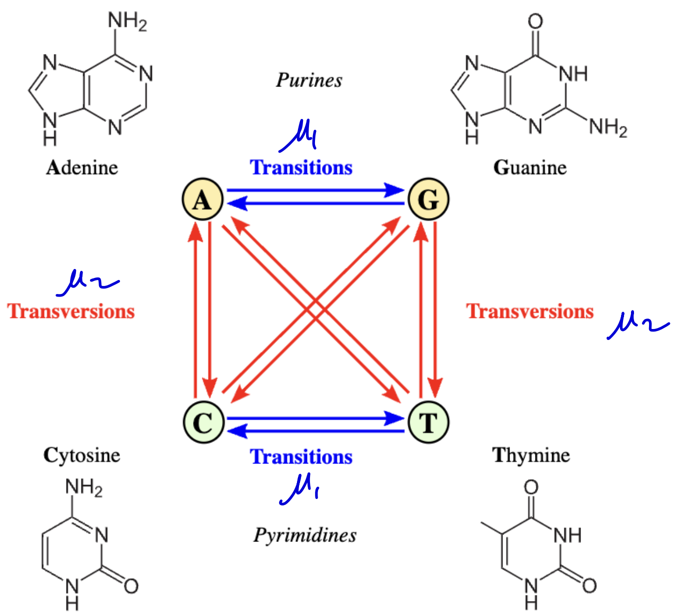
0.5

time-heterogeneous

[1] (c) The data sown is naturally modeled as a trajectory of a Poisson (a) process which is an example of a Discrete (b) -space Continuous (c) -time stochastic process.

0.25 0.25

4. A researcher is tracking the occurrence of two different types of mutations between the four DNA nucleotide (A,C,G, & T): transitions (shown with the red/solid arrows below) versus transversions (shown with the blue/dashed arrows below).



[2] (a) Transitions occur much more often than transversions. Suppose that the probability that a transition mutation occurs in a given generation is $\mu_1 = 0.2 \frac{\text{mutations}}{\text{generation}}$ and the probability that a transversion occurs is $\mu_2 = 0.1 \frac{\text{mutations}}{\text{generation}}$. Propose a transition probability matrix describing the nucleotide (A,C,G,T) state at a particular site in the genome in generation $n \in \mathbb{N}$.

off diagonal
$\mu_1 = 0.2$ 1pt
$\mu_2 = 0.1$
Diagonal
$1 - 2\mu_2 - \mu_1 = 0.6$ 1pt

P =

	A	C	G	T
A	0.6	0.1	0.2	0.1
C	0.1	0.6	0.1	0.2
G	0.2	0.1	0.6	0.1
T	0.1	0.2	0.1	0.6

[2] (b) A nucleotide is currently an Adenine (A), what is the probability that the next mutation is a transition?

Total Rate = 1 − 0.6 = 0.4 (1)

Transition = 0.2

$$\frac{0.2}{0.4} = \frac{1}{2} \quad (1)$$

[2] (c) Give the system of equations that could be solved for the stationary distribution of this markov chain?

(1pt) Remove 1

$$\begin{cases} \pi_A = 0.6\,\pi_A + 0.1\,\pi_C + 0.2\,\pi_G + 0.1\,\pi_T \\ \pi_C = 0.1\,\pi_A + 0.6\,\pi_C + 0.1\,\pi_G + 0.2\,\pi_T \\ \pi_G = 0.2\,\pi_A + 0.1\,\pi_C + 0.6\,\pi_G + 0.1\,\pi_T \\ \pi_T = 0.1\,\pi_A + 0.2\,\pi_C + 0.1\,\pi_G + 0.6\,\pi_T \end{cases}$$

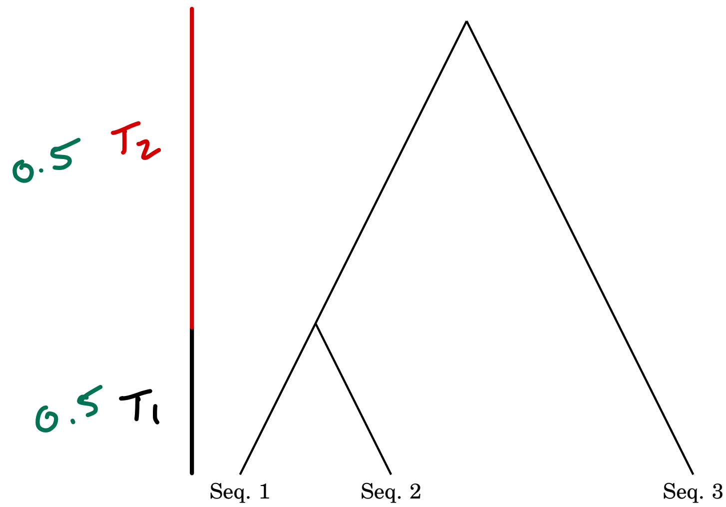
$$\pi_A + \pi_C + \pi_G + \pi_T = 1 \quad (1\text{pt})$$

[1] (d) Are there any absorbing states in this process? If so what are they? If not what kind of states are they?

No. 0.5 Recurrent 0.5

Math 468, Spring 2023
Midterm

5. Consider the genealogy and associated sequences shown below.



Seq 1	G	T	T	A	T	C	C	T
Seq 2	G	A	T	A	T	C	G	T
Seq 3	G	A	T	C	T	G	A	C

[2] (a) Label the coalescent times in the figure what is the expected value of each (measured in coalescent time units)?

$$E[T_1] = \frac{1}{\binom{3}{2}} = 0.33 \quad (0.5)$$

$$E[T_2] = \frac{1}{\binom{2}{2}} = 1 \quad (0.5)$$

[2] (b) What is the probability that the first coalescent event occurs is less then 50 Wright-Fisher generations if the effective population size is $N_e = 500$?

$$Pr(T = t) = \binom{3}{2} e^{-\binom{3}{2}t}$$

$$Pr(T < t) = 1 - e^{-\binom{3}{2}t} \quad (1pt)$$

$$50 \text{ gen} \mid 500 \text{ ind} = \frac{50}{500} = 0.1 \text{ Coal gen. } (1pt)$$

$$P(T < 0.1) = 1 - e^{-3 \cdot 0.1}$$

[1] (c) How many segregating sites are there in the data?

5

[1] (d) What is the observed number of pairwise differences between sequence 1 and sequence 2, $\pi_{1,2}$? What is the observed average number of pairwise differences, π?

$\pi_{12} = 2$ (0.5)

$\pi_{13} = 5$

$\pi_{23} = 4$

$\frac{2+5+4}{3} = \frac{11}{3}$ (0.5)