Math 211: Statistics

March 12, 2014

Spring 2014, EXAM 1

Instructions: There are four questions in this exam. Please inspect the exam and make sure you have all four questions. You may only use your calculator and A4-size **one page** of help sheet. Do all your work on the paper provided.

Remember: ***You must show your work to get proper credit.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME:____

1	/25
2	/30
3	/30
4	/25
Total:	/110

(10 points bonus)

Question 1: (25 Points) The four nucleotides adenine (A), guanine (G), cytosine (C), and uracil (U) play a key role in the formation of amino acids in the cell. A nucleotide chain is a sequence of 3 nucleotides. Assume A,G,C or U can appear any number of times (up to 3, of course) in a nucleotide chain and that all sequences are possible.

a) How many elements does the sample space of nucleotide chains have?

$\underline{4} \times \underline{4} \times \underline{4} = 4^3 = 64$ elements

b) Find the probability that a nucleotide chain has no U in it. → so 3 choices left

$\underline{3} \times \underline{3} \times \underline{3} \Rightarrow 3^3 \Rightarrow P(\text{no } U) = \frac{3^3}{4^3} = \frac{27}{64} = P(B)$

c) Find the probability that a nucleotide chain has $\overbrace{\text{at least one G}}^{A}$ given that $\underbrace{\text{it has no U in it}}_{B}$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A\cap B) = P(B) - P(A^c \cap B) = \frac{27}{64} - \frac{2^3}{64} = \frac{19}{64}$$

$\boxed{A^c \cap B : \text{no } G \text{ and no } U}$

$$\Rightarrow P(A|B) = \frac{19/64}{27/64} = \frac{19}{27}$$

d) Suppose a nucleotide chain is selected at random. Let X denote the number of U's in the selected nucleotide chain. Find the probability distribution of X.

$$\begin{cases} P(X=0) = \frac{27}{64} \text{ from part b)} \\ P(X=1) = \frac{(1\times 3\times 3)\times 3}{64} = \frac{27}{64} \\ P(X=2) = \frac{(1\times 1\times 3)\times 3}{64} = \frac{9}{64} \end{cases} \qquad P(X=3) = \frac{1}{64}$$

$\underline{U}\ \underline{\overset{A}{\underset{C}{G}}}\ \underline{\overset{A}{\underset{C}{G}}}$

$1 \times 3 \times 3$

similarly

or $\underbrace{U \text{ at any other position}}_{3 \text{ possibilities}}$

In fact, this is a Binomial r.v., at each trial $P(U) = \frac{1}{4}$ and $n=3$.

So $P(X=x) = \binom{3}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x} \quad x = 0,1,2,3.$

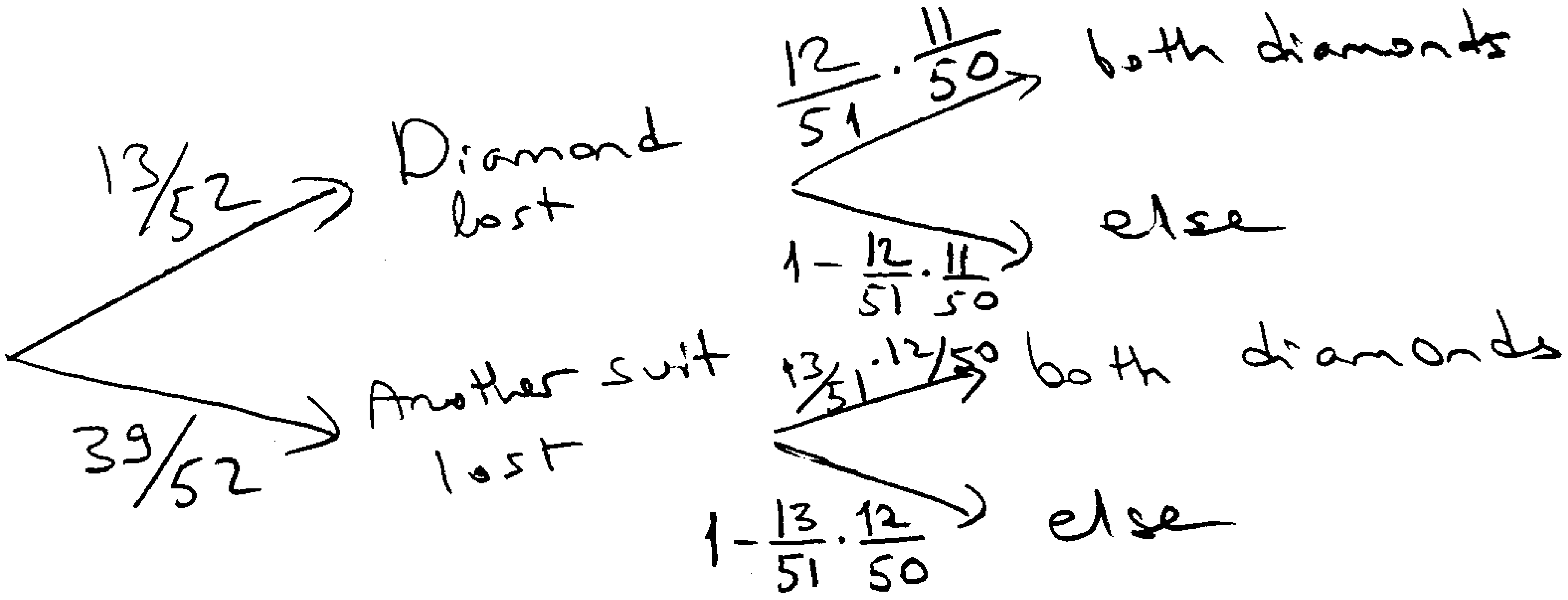
e) Let Y be 0 if there is a nucleotide appearing more than once in a nucleotide chain, and let it be 1 if none of the nucleotides appear repeatedly. Find $P(X = 1, Y = 0)$.

$P(X=1, Y=0) = P(\text{exactly one U and a nucleotide appears more than once})$

That is, $\underbrace{\underline{U}\ \underline{\overset{C}{\underset{A}{G}}}\ \underline{\overset{C}{\underset{A}{G}}}}_{3 \text{ possibilities}}$ and $\underbrace{U \text{ in different positions}}_{3 \text{ possibilities}} \times = 9$

$$\Rightarrow P(X=1, Y=0) = \frac{9}{64}$$

Question 2: (30 Points) a) A card from a deck of 52 cards is lost. From the remaining cards of the deck, two cards are drawn and are found to be both diamonds. Find the probability of the lost card being a diamond.



$$P(\text{lost card diamond} \mid \text{both diamonds}) = \frac{13/52 \cdot 12/51 \cdot 11/50}{13/52 \cdot 12/51 \cdot 11/50 + 39/52 \cdot 13/51 \cdot 12/50}$$

$$= 11/50$$

b) In a course, there are four freshman boys, six freshman girls, and six sophomore boys. How many sophomore girls must be present if gender and class are to be independent when a student is selected at random?

FB	FG	SB	SG	Total
4	6	6	n	n+4+6+6=16+n

$$P(S) \cdot P(B) = P(S \cap B)$$

$$\frac{n+6}{16+n} \cdot \frac{10}{16+n} = \frac{6}{16+n}$$

$$\Rightarrow (n+6)\,10 = 96 + 6n$$

$$4n = 36$$

$$n = 9$$

Question 3: (30 Points) Suppose that Y is an exponential random variable, that is, $f_Y(y) = \lambda e^{-\lambda y}, y \geq 0$.

a) Show that the expectation of Y is $1/\lambda$.

$$E(Y) = \int_0^\infty y \lambda e^{-\lambda y}\, dy$$

$$= \lambda \left[\cancel{\frac{e^{-\lambda y}}{-\lambda} y \Big|_0^\infty} \to 0 - \int_0^\infty \frac{e^{-\lambda y}}{-\lambda} dy \right]$$

$$= \frac{e^{-\lambda y}}{-\lambda} \Big|_0^\infty = \frac{1}{\lambda}$$

b) Derive the cumulative distribution function of Y.

$$F_Y(y) = \int_0^y \lambda e^{-\lambda y}\, dy = \lambda \frac{e^{-\lambda y}}{-\lambda} \Big|_0^y = 1 - e^{-\lambda y},$$

$$y \geq 0$$

c) If $\lambda = 0.4$, find $P(Y > 2.7)$. Suppose Y denotes the waiting time (in minutes) for the next car in metro. If five independent passengers are considered, what is the probability that exactly three of them wait at least 2.7 minutes?

$$P(Y>2.7) = 1 - F_Y(2.7) = 1 - \left(1 - e^{-(0.4)(2.7)}\right)$$

$$\cong 0.34$$

$X \sim Bin(5, 0.34)$

$$P(X=3) = \binom{5}{3} (0.34)^3 (0.66)^2$$

$$\cong 10 (0.039)(0.43) \cong 0.17$$

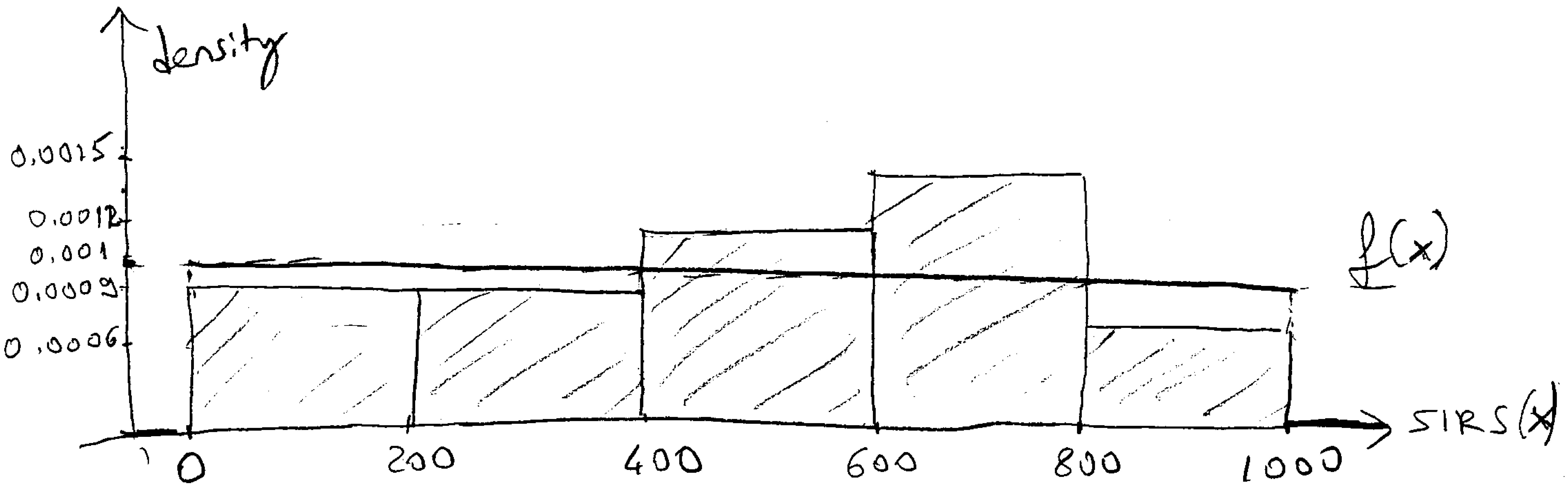
Question 4: (25 Points) Medical conditions are compared on a "seriousness of illness rating scale" (SIRS) (that is, SIRS quantifies how serious a medical condition is). SIRS scores are given below for seventeen medical conditions such as high blood pressure, diabetes, alcoholism:

21, 173, 174, 204, 312, 393, 454, 468, 500, 520, 621, 636, 688, 733, 776, 824, 997

a) Form a frequency table by grouping the data in class intervals of size 200.

Class Inter.	0-200	200-400	400-600	600-800	800-1000
Frequency	3	3	4	5	2
Rel. freq.	3/17	3/17	4/17	5/17	2/17
Density	$3/17 \cdot \frac{1}{200}$	$3/17 \cdot 1/200$	$4/17 \cdot 1/200$	$5/17 \cdot 1/200$	$2/17 \cdot 1/200$

b) Draw a density scale histogram.



c) Write down the uniform probability density function (pdf) that models this data set, and plot it on the density scale histogram.

$$\text{Unif}(0, 1000) \Rightarrow f(x) = \frac{1}{1000} \qquad 0 \leq x \leq 1000$$

d) Find the cdf for the uniform pdf that you used in part c). Make sure to give the values of $F(x)$ for all $x \in R$.

(If you have not answered part c), use uniform distribution on the interval [0,1] for part d)).

$$F(x) = \int_0^x \frac{1}{1000}\,dx \quad \text{if } 0 \leq x \leq 1000$$

$$= \frac{x}{1000} \quad \text{if } 0 \leq x \leq 1000$$

$$F(x) = 0 \quad \text{if } x < 0$$

$$F(x) = 1 \quad \text{if } x > 1000$$