Name: ____

ID: ____

On your desk you should have only the exam paper, writing tools, and the cheat-sheet. The cheat-sheet is one page handwritten on both sides.

The exams are color coded. Your exam should have different color than that of your neighbours to the left, right and in front.

There are 10 questions in this exam, totalling 100 points and an extra-credit question, worth 10 points. The final score is determined by summing all the points and taking the min of the sum and 100.

Be clear and concise. Write your answers **in the space provided** after each question. To the degree possible, write your answers as expressions. If you provide only the final numerical answer it would be hard for us to assign you partial credit. Similarly, clearly written explanations written after the final answer will make it possible for us to give you partial credit if your final answer is incorrect.

Do your work on the back side of the pages and then, when you are confident of the answer, copy it, including explanations, neatly into the space below the question.

1	10pt.
2	8pt.
3	7pt.
4	15pt.
5	10pt.
6	10pt.
7	10pt.
8	10pt.
9	10pt.
10	10pt.
EC	10pt.
total	

1. (10 pts)

How many strings of 3 lower case letters are there that contain at least 2 of the letters in the range a-g inclusive?

Hint: look at the complement: how many strings contain no letters in a-g, and how many strings contain exactly one letter in a-g?

____

2. (8 pts)

Suppose a pedestrian crosses a busy street either on or off the crosswalk.Suppose that the probability she uses the cross-walk is 90%. Furthermore, suppose that the probability of getting hit by a car is 10^{-7} when using the cross-walk and 10^{-4} when not using the cross-walk.

Suppose the pedestrian crossed the street and was hit by a car, what is theprobability that she was using the cross-walk? ____________________

Suppose the pedestrian crossed the street and was not hit by a car, what is theprobability that she was using the cross-walk? ____________________

3. (7 pts)

You are giving candy to 6 kids. The candy consists of 6 Twix and 7 Milky-ways. Each kids gets at least one Twix. How many ways are there to distribute the candy among the kids?

____

4. (15 pts)

For the following joint probability matrices, fill in the missing values and answer the questions.

	X=1	X=2	X=3	Marginal over Y
Y=1		0	1/8	
Y=2	0		0	1/2
Y=3	1/8	0	1/8	1/4
Marginal over X	1/4	1/2	1/4	*

- Are X and Y dependent? ____________________
- Are X and Y correlated? ____________________
- $P(Y = 3|X = 1) =$ ____________________

	X=1	X=2	X=3	Marginal over Y
Y=1		0	1/5	1/5
Y=2	1/5		0	2/5
Y=3	0	2/5		2/5
Marginal over X	1/5	3/5	1/5	*

- Are X and Y dependent? ____________________
- Are X and Y correlated? ____________________
- $P(X = 1|Y = 2) =$ ____________________

	X=1	X=2	X=3	Marginal over Y
Y=1	1/6	1/12	1/12	1/3
Y=2	1/6	1/12	1/12	1/3
Y=3		1/12	1/12	
Marginal over X	1/2		1/4	*

- Are X and Y dependent? ____________________
- Are X and Y correlated? ____________________
- $P(Y = 2|X = 3) =$ ____________________

5. (10 pts)

You are given a card deck of 44 cards, consisting all combinations of 11 ranks and 4 suits. Suppose you draw a hand of 5 cards out of this deck, what is the probability that it will be a *straight* ?

Recall that a *straight* refers to five cards whose ranks are in sequence, but whose suits are **not all the same** (otherwise it is called a straight flush). The 1 can be used either as a low card or a high card when forming a straight. For example, both (1,2,3,4,5) and (8,9,10,11,1) are valid straight ranks, but not (9,10,11,1,2).

1. The number of possibilities for the ranks of a straight is ________________ .

2. The suits can be anything other than all equal, so the number of possibilities for the suits of a straight is ________________ .

3. Thus the number of straight hands is ________________ .

4. Finally, the probability that a randomly chosen hand is a straight is ________________ .

6. (10 pts)

You are given a coin that you assume is a fair coin, and you are interested in the probability that the fraction of "heads" in the sequence is at least 0.4

a. Suppose the coin is flipped n=100 times

- Write an exact expression for the probability that the number of heads is equal to $0.4 \times 100 = 40$. ____________________
- What Z-score corresponds to getting 40 heads? ____________________
- What is the central limit theorem approximation for the probability that the number of heads is 40 or higher assuming that the coin is fair? ____________________

b. Suppose that instead of n=100 times we flip the coin n=10000 times.

- What Z-score corresponds to getting $0.4 \times 10000 = 4000$ heads? ____________________
- What is the central limit theorem approximation for the probability that the number of heads is 4000 or higher assuming that the coin is fair? ____________________

a. As we increase the number of coin flips n, does the probability of the event **"the number of heads is larger than 0.4**$\times n$**"** increase,decrease or stay the same? ____________________

7. (10 pts)

Suppose a web-server receives, on average, 60 requests per minute. Assume that the requests are independent and therefore:

1. The number of requests in any given minute is distributed according to a Poisson distribution and
2. The time between two consecutive requests is distributed according to an Exponential distribution.

- ◦ What is the probability that it will not receive any request within the first 3 seconds of some particular minute? ____________________
- ◦ What is the probability that it will receive exactly 3 requests during the same period of time? ____________________
- ◦ What is the probability that the time between the first and the second requests received in a minute would be smaller than 5 seconds? ____________________

8. (10 pts)

Suppose $X_1, ..., X_{92}$ are independent random variables where $X_i \in \{0, 1\}$ and $P(X_i = 1) = 0.3$ for $i = 1, ..., n$.

Define a random variable $Y = \sum_{i=1}^{92} (-1)^{i+1} X_i = X_1 - X_2 + X_3 - X_4 + ... - X_{92}$

What is $E[Y]$? ____________________

What is $Var[Y]$? ____________________

Using Chebyshev's inequality, give an upper bound on $P(|Y| > 30)$? ____________________

9. (10 pts)

We want to prove that seat-belts save lives. In other words, our goal is to reject the null hypothesis stating that seat-belts do not help.

Suppose first that we know that, in general, the probability of a fatalityin a car accident is $q = 1\%$.

We examine $n = 1000$ randomly selected records of accidents where the driverwore a seat belt. We find that $k = 7$ of these accidents were fatal.

With what confidence can we conclude that claim iscorrect?X_i = i^{th} accident was fatal.

The null hypothesis is that seat belts make no difference.$Pr(X_i = 1) = q = 0.01$. Alternative hypothesis, $Pr(X_i = 1) < q$.

The total number of fatal accident under the null hypothesis $S = \sum_i X_i$.

$E(S) =$ ____________________

$var(S) =$ ____________________ $std(S) =$ ____________________

$k = 7$ is approximately one standard deviation smaller than the mean, in other words, the z-score is roughly (write an integer number) ____________________

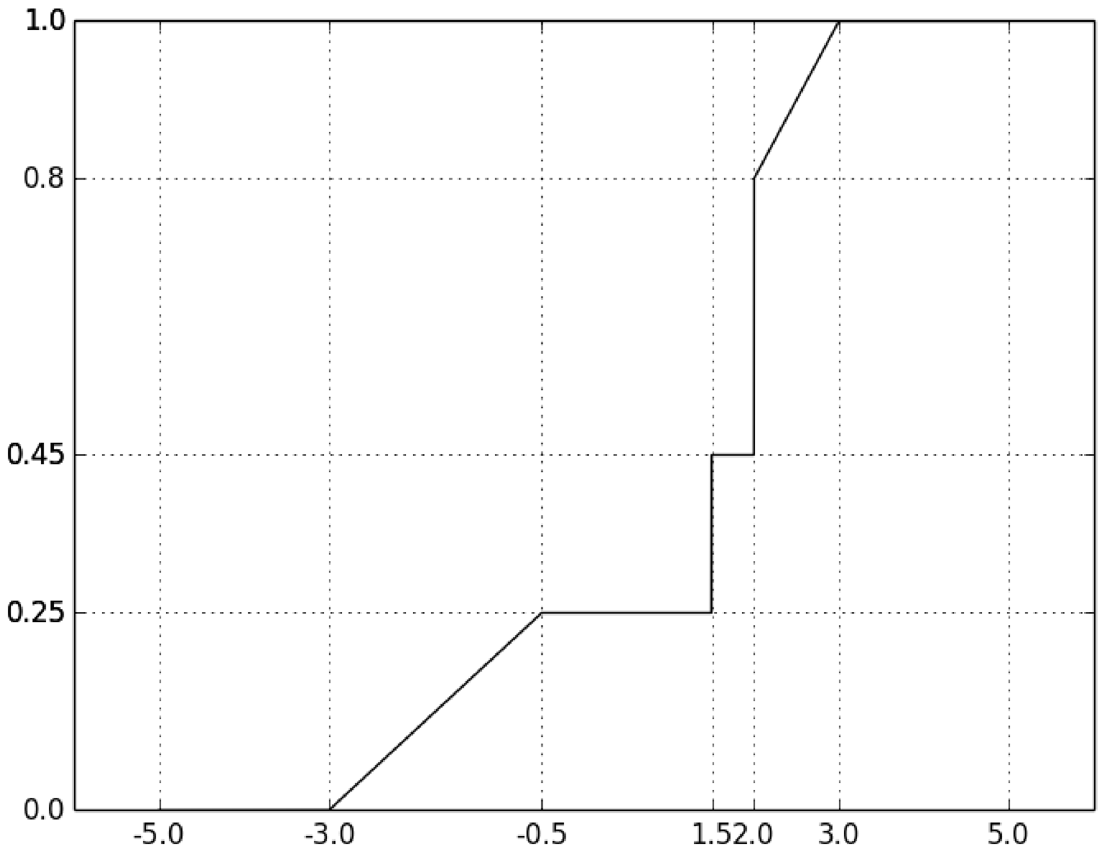
Using this approximation, the p**-value** associated with $k = 7$ is ____________________ .

Complete the sentence: the p-value ("is"/"is not") ____________________ a **random variable**.

10. (10 pts)

Below is the CDF of a mixture distribution with point mass and uniform components.

Component weights take on multiples of 0.05 and they need to sum to one.



Identify the component distributions:

- Uniform component on the interval (________________ , ________________). Its component weight is ________________
- Uniform component on the interval (________________ , ________________). Its component weight is ________________
- Point mass on ________________ . Its component weight is ________________
- Point mass on ________________ . Its component weight is ________________

11. (10 pts)

(Extra Credit)

Suppose you have an algorithm A for your problem that always returns the correct answer, but takes different amounts of time each time it runs. Let X_n be the random variable giving the time algorithm A takes to complete for input of size n. Time can't be negative, so $X \geq 0$.
We call A a *Las Vegas Algorithm* if for any input size n there is a $T(n)$ so that $\mathbb{E}(X_n) = T(n)$.

- Side note: it isn't always the case that a random variable has finite expectation, even the values it can take on are finite. The assumption that there is some $T(n)$ for any n is a non-trivial assumption.

Let's say we have algorithm that determines whether any integer is prime. For an integer input that takes up n bits, the algorithm takes n seconds to run with probability $1/2$. With probability $1/2$ the algorithm takes 1 second to run. What is $T(n)$, in seconds? __________________ .

Let's say you have some Las Vegas algorithm A that runs in *expected time* $T(n)$ for problem size n. What you would prefer, however, is an algorithm that *always* finishes in time $O(T(n))$, but may have up to a 5% probability of returning the wrong answer. We will construct an algorithm A' from A that satisfies these properties.

Recall Markov's inequality: for some random variable $Y \geq 0$, $P(Y >= a) \leq \mathbb{E}(Y)/a$.
Fixing some n, apply Markov's inequality to get an upper bound on $P(X_n >= cT(n))$: __________________ .
What is c such that $P(X_n >= cT(n)) <= 0.05$? __________________

Thus an algorithm A' is as follows:

- Run A until time __________________ T(n). If A has completed, return the correct value.
- Else, return a random value.

This type of algorithm we have, where the algorithm completes in deterministic time $Q(n)$ but is correct with some probability is called a *Monte Carlo Algorithm*.

Recap:

- For "Las Vegas" algorithms, uncertainty is in the algorithm *runtime*. The algorithm is always correct.
- For "Monte Carlo" algorithms, uncertainty is in the algorithm *correctness*. The algorithm completes in deterministic time $T(n)$.