Midterm 1

Last Name	First Name	SID

- You have 10 minutes to read the exam and 90 minutes to complete this exam.
- The maximum you can score is 120, but 100 points is considered perfect.
- The exam is not open book, but you are allowed to consult the cheat sheet that we provide. No calculators or phones.
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- Show all work to get any partial credit.
- Take into account the points that may be earned for each problem when splitting your time between the problems.

Problem	points earned	out of
Problem 1		30
Problem 2		20
Problem 3		10
Problem 4		20
Problem 5		20
Problem 6		20
Total		100 (+20)

Problem 1: Answer these questions briefly but clearly. [30]

(a) [**6**] Let $X \sim \mathcal{N}(0,1)$. Compute $\mathbb{E}[|X|]$.

(b) [**6**] Suppose that you are in a Second-Price Auction with n other bidders, and all other bidders draw their valuations uniformly and independently from the interval $[0,1]$. If your valuation is v, and everyone including you bids their own valuation, what is your expected profit from the auction? (Remember that if you don't win the auction, you don't pay anything.)

(c) [**6**] Recall the soliton distribution we used for encoding messages in the fountain codes lab. In this setting, n data chunks were encoded into n packets, which are XORs of d many randomly selected data chunks, where d is drawn from a degree distribution $p(\cdot)$. Assuming a perfect channel, where the probability of a packet being erased is 0, **show that the probability that this scheme fails to decode the original message,** $\mathbb{P}(\textbf{failure}) \geq \mathrm{e}^{-1}$ **as** $n \to \infty$**.**

The Taylor series $\mathrm{e}^x = \lim_{n\to\infty}(1+\frac{x}{n})^n$ may come in handy.

$$p(d) = \begin{cases} 0, & d \leq 0 \text{ or } d > n \\ \frac{1}{n}, & d = 1 \\ \frac{1}{d(d-1)}, & 1 \leq d \leq n \end{cases}$$

(d) [6] Tom Brady just lost the Super Bowl, and he needs to validate that he is strong enough to keep playing football. He decides to throw a football as far as he can each day, and keep track of the farthest throw he has recorded so far; we will call this farthest throw his personal best. The distance that he throws the football is drawn from a continuous non-negative distribution and all throws are i.i.d. **What is the expected number of times, $\mathbb{E}[T]$, that his personal best changes over n throws?**

Assume that on day one, his personal best is distance 0 yards.

(e) [6]

We consider a bag with red and blue balls. The number of red balls is Poisson distributed with parameter λ_{r}, and the number of blue balls is independently Poisson distributed with parameter λ_{b}. Conditioned on there being n (greater than 1) balls in the bag, what is the distribution of the number of blue balls? (Leave your answer in terms of n, λ_{r}, and λ_{b}.)

Problem 2: Bounds [20]

(a) [**10**] We have a random walk over the integers starting at zero. Each time we either move left or right with equal probability. Let S_n be a random variable which is equal to the integer that we lie on at time n. So $S_0 = 0$ and $S_n = Y_1 + \cdots + Y_n$, where each Y_i is either $+1$ or -1 with probability $1/2$, independently.

Show that

$$\mathbb{P}(|S_n| \geq t) \leq 2\mathrm{e}^{-\frac{t^2}{2n}}, \quad \text{for any } t \geq 0.$$

Hint: Recall that in homework 4 you showed that if $X_1, \ldots, X_n$ are independent Bernoulli(q), then

$$\mathbb{P}(|X_1 + \cdots + X_n - nq| \geq \epsilon) \leq 2\mathrm{e}^{-\frac{2\epsilon^2}{n}}, \quad \text{for any } \epsilon \geq 0.$$

(b) [**10**] We have seen that the Markov bound can be quite loose. However, this need not always be true. For a given positive integer k, describe a random variable X that assumes only non-negative values such that: $\mathbb{P}(X \geq k\,\mathbb{E}(X)) = 1/k$.

Problem 3: Entropy [10]

Recall the definitions of entropy and joint entropy for discrete random variables:

$$
\begin{aligned}
H(X) &\triangleq \mathbb{E}[-\log_2 p_X(X)] = -\sum_x p_X(x)\log_2 p_X(x),\\
H(X,Y) &\triangleq \mathbb{E}[-\log_2 p_{X,Y}(X,Y)] = -\sum_x\sum_y p_{X,Y}(x,y)\log_2 p_{X,Y}(x,y).
\end{aligned}
$$

(a) [5] Let $U \sim \text{Uniform}\{1,2,\ldots,n\}$. Give a closed form expression for $H(U)$.

(b) [5] Assume that X, Y are independent. Express $H(X,Y)$ in terms of $H(X)$ and $H(Y)$.

Problem 4: Transformations of Random Variables [20]

Consider X, Y i.i.d. Uniform$[0,1]$. Let $Z = \ln \frac{X}{Y}$.

(a) [**5**] Show that $-\ln Y$ is exponentially distributed with parameter 1.

(b) [**5**] Find the moment generating function of Z, M_Z.

(c) [**5**] Find $\operatorname{var} Z$.

(d) [**5**] Find the PDF of Z, f_Z.

Problem 5: Valentine's Day [20]

The joint density of (X, Y) is uniform on the shaded region in Figure 1. Mathematically, the shaded region consists of two half-circles (each of radius one) centered at $(-1, 0)$ and $(1, 0)$, along with a triangle in the lower half-plane.

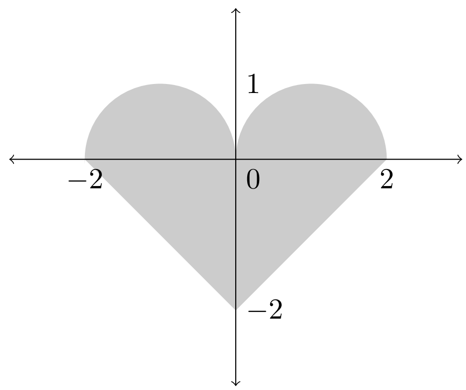

Figure 1: Joint density of (X, Y).

(a) **[7]** Find the value of the joint density in the shaded region.

(b) **[8]** Find the marginal density of Y.

(c) **[5]** Find $\text{cov}(X, Y)$. (Don't handwave—explain your answer carefully.)

Problem 6: Waiting in Line [20]

You are first in line to be served by either of 2 servers, Alice and Bob, who are busy with their current customers. You will be served as soon as the first server finishes. The service times of Alice and Bob are independently and exponentially distributed with (positive) rates a and b respectively, i.e. Alice's distribution is Exponential(a), and Bob's is Exponential(b).

(a) [5] Find the probability that Alice will be your next server.

(b) [5] True or False? (you must prove your answer): The probability that you will be the last to be served (among you and the two current customers) is less than a half if and only if a is not equal to b.

(c) [10] What is the distribution of your wait time? Your answer should not include integrals.