Read directions carefully and show all your work. Particularly, define the events and random variables you are considering. Partial credit will be assigned based upon the correctness, completeness and clarity of your answers. Correct answers without proper justification will not receive full credit.

The exam is closed book, closed notes. Calculators are permitted but not needed to answer questions.

APPENDIX

Probability Mass Functions of Common Discrete Distributions

- Discrete uniform distribution: $p_X(x_i) = 1/n$, for $x_i \in X(\Omega) = \{x_1, \ldots, x_n\}$
- Binomial distribution with parameters $n$ and $p$: $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, \ldots, n$
- Geometric distribution with parameter $p$: $p_X(x) = (1-p)^{x-1} p$, for $x \in \mathbb{N}, x \geq 1$
- Poisson distribution with parameter $\lambda$: $p_X(x) = e^{-\lambda} \lambda^x / x!$, for $x \in \mathbb{N}$

Probability Density Functions of Common Continuous Distributions

- Continuous uniform distribution with parameters $a$ and $b$, $(a < b)$: $f_X(x) = 1/(b-a) \mathbb{1}_{[a,b]}(x)$
- Normal distribution with parameters $\mu$ and $\sigma^2$: $f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\}$
- Exponential distribution with parameter $\lambda$: $f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$
Problem 1. [7 points] The amount of money spent by a customer at a bakery is modeled by a normal distribution with mean $10 and standard deviation $2.

(a) [2 points] What is the probability that a randomly selected customer spends less than $7?

Answer. Let $X$ be the amount of money spent by a customer. The probability that a randomly selected customer spends less than $7$ is

$$
P(X < 7) = P \left( \frac{X - 10}{2} < \frac{7 - 10}{2} \right) = \Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332.
$$

(b) [3 points] The owner wants to know the amounts of money that the middle 60% of the customers spend. Find the upper and lower amounts of spending of this group of customers.

Answer. We denote $x_{\text{upper}}$ and $x_{\text{lower}}$ the upper and lower amounts of spending of the group of customers. These quantiles verify

$$
P(x_{\text{lower}} \leq X \leq x_{\text{upper}}) = 0.6 \iff P \left( \frac{x_{\text{lower}} - 10}{2} \leq \frac{X - 10}{2} \leq \frac{x_{\text{upper}} - 10}{2} \right) = 0.6.
$$

Note that $(x_{\text{upper}} - 10)/2$ is the the quantile of the standard normal distribution of order 0.8 which is 0.84. Therefore, $x_{\text{upper}} \approx 2 \cdot 0.84 + 10$. Due to the symmetry of the Gaussian distribution, $x_{\text{lower}} \approx 2 \cdot (-0.84) + 10$.

(c) [2 points] The owner picks at random receipts of customers from a pile until she finds one that falls into the category described in the preceding question. What is the probability that she selects exactly 5 receipts before she finds one that corresponds to the desired group?

Answer. Let $Y$ be the number of receipts picked until the owner finds one that falls into the category of the “middle” 60%. Then $Y$ follows a geometric distribution with parameter 0.6. Hence, the desired probability is

$$
P(Y = 5) = (1 - 0.6)^4 \cdot 0.6.
$$

You are given the table below for the cumulative distribution function of the standard normal distribution.

<table>
<thead>
<tr>
<th>$z$</th>
<th>0.80</th>
<th>0.81</th>
<th>0.82</th>
<th>0.83</th>
<th>0.84</th>
<th>0.85</th>
<th>0.86</th>
<th>0.87</th>
<th>0.88</th>
<th>0.89</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi(z)$</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>$z$</td>
<td>1.50</td>
<td>1.51</td>
<td>1.52</td>
<td>1.53</td>
<td>1.54</td>
<td>1.55</td>
<td>1.56</td>
<td>1.57</td>
<td>1.58</td>
<td>1.59</td>
</tr>
<tr>
<td>$\Phi(z)$</td>
<td>0.9332</td>
<td>0.9345</td>
<td>0.9357</td>
<td>0.9370</td>
<td>0.9382</td>
<td>0.9394</td>
<td>0.9406</td>
<td>0.9418</td>
<td>0.9429</td>
<td>0.9441</td>
</tr>
</tbody>
</table>
Problem 2. [5 points] A continuous random variable $X$ has a probability density function given by
\[ f_X(x) = kx(a-x) \mathbb{I}_{[0,3]}(x) \]
where $k$ and $a$ are positive constants.

(a) [1 point] Explain why $a \geq 3$.

Answer. Since $f_X$ is a probability density function, $kx(a-x) \geq 0$ for $x \in [0,3]$. In particular, for $x = 3$, this translates into
\[ 3k(a-3) \geq 0 \Rightarrow a-3 \geq 0. \]

(b) [2 points] Show that $k = \frac{2}{9(a-2)}$.

Answer. $f_X$ integrates to 1:
\[ \int_0^3 kx(a-x) \, dx = 1 \Leftrightarrow k \left[ \frac{x^2}{2} - \frac{x^3}{3} \right]_0^3 = 1 \Leftrightarrow k = \frac{1}{9(a/2-1)}. \]

(c) [2 points] Given that $E[X] = 1.75$, show that $a = 4$.

Answer. We have that
\[ E[X] = 1.75 \Leftrightarrow \int_0^3 kx^2(a-x) \, dx = 1.75 \Leftrightarrow k \left[ \frac{x^3}{3} - \frac{x^4}{4} \right]_0^3 = \frac{7}{4} \Leftrightarrow \frac{2}{9(a-2)} \left( \frac{9a}{4} - \frac{81}{4} \right) = \frac{7}{4}. \]
**Problem 3.** [9 points] The continuous random variable $S$ represents the time in hours per week that people spend playing soccer. The probability density function of $S$ is

$$f_S(t) = \begin{cases} 
(1/9) \cdot (t^2 - 2t + 2) & 0 < t \leq 3 \\
1/3 & 3 < t \leq 4 \\
0 & \text{otherwise}
\end{cases}$$

(a) [3 points] Find the cumulative distribution function $F_S$ of $S$.

**Answer.** For $t \in (0, 3]$,

$$F_S(t) = F_S(0) + \int_0^t \frac{1}{9} (x^2 - 2x + 2) \, dx = \frac{1}{9} \left( \frac{t^3}{3} - t^2 + 2t \right).$$

For $t \in (3, 4]$,

$$F_S(t) = F_S(3) + \int_3^t \frac{1}{3} \, dx = \frac{1}{9} \left( \frac{3^3}{3} - 3^2 + 2 \cdot 3 \right) + \frac{t - 3}{3} = \frac{t - 1}{3}.$$

In a nutshell,

$$F_S(t) = \begin{cases} 
0 & t < 0 \\
\frac{1}{9} \left( \frac{t^3}{3} - t^2 + 2t \right) & 0 < t \leq 3 \\
\frac{t - 1}{3} & 3 < t \leq 4 \\
1 & t > 4
\end{cases}$$

(b) [2 points] Show that the proportion of the population that spends more than 3 hours playing soccer is $1/3$.

**Answer.** The proportion of the population that spends more than 3 hours playing soccer is given by

$$\Pr(S > 3) = 1 - F_S(3) = 1 - \frac{2}{3} = \frac{1}{3}.$$

(c) [4 points] Actually, the above model is relevant for amateurs. Instead, an exponential distribution is more appropriate for professional soccer players. According to a study, there is 0.1% of professional players in the population and they spend an average of 20 hours a week training/playing soccer. A person is selected at random in the population. Given that the person spends more than 3 hours a week playing soccer, find the probability that he/she is a professional player.

**Answer.** Let $T$ be the time in hours that a person spends playing soccer. We denote $P$ the event that a person is a professional player. Note that we are given that $\Pr(P) = 0.001$. Then, for professional players, $T|P$ follows an exponential distribution with parameter $1/20$. On the other other hand, for amateurs, $T|P^c$ follows the model given by the probability density function $f_S$. Hence, the conditional statement translates into the following computation:

$$\Pr(P|T > 3) = \frac{\Pr(T > 3|P)\Pr(P)}{\Pr(T > 3|P)\Pr(P) + \Pr(T > 3|P^c)\Pr(P^c)} \quad \text{(Bayes' Formula)}$$

$$= \frac{\int_3^{\infty} (1/20) \cdot e^{-x/20} \, dx \cdot 0.001}{\int_3^{\infty} (1/20) \cdot e^{-x/20} \, dx \cdot 0.001 + (1/3) \cdot 0.999}$$

$$= \frac{e^{-3/20} \cdot 0.001}{e^{-3/20} \cdot 0.001 + (1/3) \cdot 0.999}$$
Problem 4. [5 points] Let $X$ be a continuous random variable. We say that the distribution of $X$ is symmetric about $m \in \mathbb{R}$ if and only if there exists a value $m$ such that

$$f_X(m - x) = f_X(m + x), \quad \text{for all } x \in \mathbb{R} \tag{1}$$

where $f_X$ is the probability density function of $X$.

(a) [2 points] Give an example of symmetric distribution among the common distributions (see Appendix). Specify the parameter(s) of the distribution and the corresponding number $m$ in Eq. (1). You do not need to justify your answer for this question.

Answer. The continuous uniform distribution $U(-1,1)$ is symmetric about $m = 0$. Another example is the normal distribution with parameters $\mu$ and $\sigma^2$ that is symmetric about $m = \mu$.

(b) [3 points] Assume that the distribution of $X$ is symmetric about 0, that is $f_X(-x) = f_X(x)$, for all $x \in \mathbb{R}$. Show that $F_X(x) + F_X(-x) = 1$, where $F_X$ is the cumulative distribution function of $X$.

Answer. By definition, $F_X(x) = \mathbb{P}(X \leq x)$. Therefore, it is sufficient to prove that $F_X(-x) = \mathbb{P}(X > x)$.

$$F_X(-x) = \int_{-\infty}^{-x} f_X(t) \, dt$$
$$= -\int_{\infty}^{x} f_X(-u) \, du \quad \text{with the change of variable } u = -t$$
$$= \int_{x}^{\infty} f_X(u) \, du \quad \text{using the symmetry } f_X(-u) = f_X(u)$$
$$= \mathbb{P}(X > x).$$