Problem 1 (Conditional Probabilities and Moments).
Consider a sample space $\Omega$ with size $|\Omega| = 100$ and events $E_1, E_2, E_3$ as in the Venn diagram below. Also assume $|E_1 \cup E_2 \cup E_3| = 96$.

(a) What is the probability of the event $A = E_1 \cap E_2 \cap E_3$?

(b) What are the conditional probabilities $P(E_1|E_3)$, $P(E_1|E_2, \bar{E}_3)$?

Define the random variables

\[ X_1 = \begin{cases} 0 & E_1 \\ 1 & \bar{E}_1 \end{cases}, \quad X_2 = \begin{cases} 0 & E_2 \\ 1 & \bar{E}_2 \end{cases}, \quad X_3 = \begin{cases} 0 & E_3 \\ 1 & \bar{E}_3 \end{cases}, \quad X_{23} = \begin{cases} 1 & E_2 \cap E_3 \\ 2 & \bar{E}_2 \cap E_3 \\ 3 & \text{otherwise} \end{cases} \]

(c) What are the probabilities $P(X_2)$, $P(X_{23})$, $P(X_{23}, X_2)$, $P(X_{23}|X_1)$ and $P(X_{23}|X_2)$?

(d) We write $X \perp Y$ if $X$ is independent with $Y$.

Without performing any computations answer whether $X_{23} \perp X_1$ and whether $X_{23} \perp X_2$.

How would you test if $X_1 \perp X_2, X_3$?

(e) What is the variance, expectation and second moment of $X_{23}$?

Problem 2 (Tail Bounds and Central Limit Theorem).
Suppose we have a computer program consisting of $n = 100$ modules of code. Let $X_i$ be the number of errors on the $i^{th}$ module of code. Suppose that $X_i \sim \text{Poisson}(1)$ are i.i.d. Let $Y = \sum_{i=1}^n X_i$ be the total number of errors.

(a) Compute Markov and Chebyshev bounds for $P(Y \geq 90)$.

(b) Use the central limit theorem to approximate $P(Y \geq 90)$.

Problem 3 (Conditional Probabilities and Bayes’ Theorem).
Consider the relationship between people watching more than 10 hours of television (TV) per week (a binary random variable $T$ set to 1 if this is the case) and people getting heart problems (a binary random variable $H$ set to 1 if this is the case). A statistician has compiled the following numbers:

- among the people with heart problems, 80% were watching more than 10 hours of television
- among the people without heart problems, 50% were watching more than 10 hours of television
- 30% of all people got heart problems

What is the probability that somebody who watches more than 10 hours of television gets heart problems?

Problem 4 (Maximum Likelihood Estimator).
You are subscribed to a service for virtual machines, each of which is assigned a serial number $x$ with $0 < x < N$, where $N$ is the total number of machines. In your last experiment you recorded the list of serial numbers $X_1, \ldots, X_n$ for your processes. You want to estimate how many machines the service has. For simplicity you may assume that the serial numbers are continuous and non-unique. To estimate the parameter $N$ from your sample $X_1, \ldots, X_n$. 

1 of ??
(a) derive the maximum likelihood estimator $\hat{N}_{\text{MLE}}$, and
(b) use the method of moments to derive estimators $\hat{N}_1$ and $\hat{N}_2$ using the 1st and 2nd moments, respectively.
(c) One estimator of the parameter $N$ is $\hat{N} = 2x$, where $x$ is a single sample from this distribution. Give an $1 - \alpha = 95\%$ confidence interval for $\hat{N}$.
   
   Hint: For an $1 - \alpha = 95\%$ confidence interval it is $P\left( N \in \text{CI}(\hat{N}) \mid N \right) = E\left[ I\left( N \in \text{CI}(\hat{N}) \right) \mid N \right] = 1 - \alpha$.
(d) The above estimator becomes $\hat{N} = \frac{2}{n} \sum_{i=1}^{n} x_i$ for $n$ samples. Give a 95% confidence interval for $n = 100$.
   
   Hint: As the number of samples increases, you may assume that $\hat{N}$ is normally distributed.

Problem 5 (Maximum Likelihood Estimator).
You take part in a curious game with 3 other players: Tom, Ellen and Marion. Before the game begins, Marion commits to a secret positive number $\alpha$, which remains constant throughout the game. On each turn, Ellen spins the top of the image on the right, and notes the lowermost position it stops at, which is a number $e \in [0, 1)$; she then whispers this to Marion. At the same time, Tom flips a coin which gives ‘tails’ with probability $\gamma$, and whispers the outcome to Marion. Finally, Marion either subtracts 1 from Ellen’s number or multiplies it with her $\alpha$, if Tom whispered ‘heads’, or ‘tails’, respectively; then she announces the result. We represent the number that Marion announces with the random variable $X \sim f_X(x; \alpha, \gamma)$. Your goal is to find out the parameters $\alpha$ and $\gamma$ of the game after $n$ turns.

(a) What is the probability density function $f_X(x; \gamma, \alpha)$?
   
   Hint: The resulting distribution is a mixture (of uniform ones).
(b) Derive the maximum likelihood estimators $\hat{\gamma}_{\text{MLE}}, \hat{\alpha}_{\text{MLE}}$.
(c) Use the method of moments to derive estimators $\hat{\gamma}_{\text{MOM}}, \hat{\alpha}_{\text{MOM}}$.

Problem 6 (Chi-Square Test). You have two bags, $A$ and $B$, each containing 5 identical coins. You want to test if the coins in each bag are fair. To do this you repeat for each bag the following experiment: you toss all coins of the bag and write the number of ‘heads’ that appear on a sheet of paper. You repeat the toss 1000 times, and then you count how many times each number appeared on the sheet for each bag, resulting in the counts in the table below. You want to show that the coins are biased.

<table>
<thead>
<tr>
<th>value</th>
<th>Coin A</th>
<th>Coin B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>162</td>
<td>141</td>
</tr>
<tr>
<td>2</td>
<td>332</td>
<td>275</td>
</tr>
<tr>
<td>3</td>
<td>290</td>
<td>322</td>
</tr>
<tr>
<td>4</td>
<td>157</td>
<td>189</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>46</td>
</tr>
</tbody>
</table>

(a) What is your null hypothesis?
(b) Write down the formula that performs the exact test, without performing any calculations.
   
   Is it computationally easy to perform your test?
(c) Write down the formula for Pearson’s $\chi^2$ test. Then perform it for both bags. Interpret the result.
   
   Hint: Review the relation between the Bernoulli, Binomial and Multinomial: each generalises the previous.