Part I: Probability Theory

1. [**6 Points**]

Four people toss, one after the other, a fair coin. We assume that the outcomes of the four tosses are independent.

(a) [**2 Points**]

Compute the probability of obtaining 'Heads' exactly once.

Solution:

We have $E = \{HTTT, THTT, TTHT, TTTH\}$ and $\mathbb{P}(E) = 4 \times (1/2)^4 = 1/4$.

(b) [**2 Points**]

Compute the probability of obtaining 'Tails' exactly twice.

Solution:

We have $\mathbb{P}(\{\text{'Tails' exactly twice}\}) = \binom{4}{2} \times \mathbb{P}(\{TTHH\}) = \frac{4!}{2!2!}\frac{1}{2^4} = \frac{3}{8}$.

(c) [**2 Points**]

Compute the probability of obtaining 'Tails' at least once.

Solution:

We have $E^c = \{HHHH\}$ and $\mathbb{P}(E) = 1 - \mathbb{P}(E^c) = 15/16$.

2. [**9 Points**]

A building has three floors. Three people get in and walk up to one of the floors. Let Ω denote the sample space, that is, the set of all possible elementary events describing which person goes to which floor.

(a) [**1 Point**]
Recall what a Laplace model is.
Solution:

Let $\Omega = \{\omega_1, \dots, \omega_N\}$ with $N = |\Omega|$. A Laplace model stipulates that $\mathbb{P}(\{\omega_i\}) = 1/N$ for all $i \in \{1, \dots, N\}$.

(b) [**2 Points**]
Compute $p(\omega) = \mathbb{P}(\{\omega\})$ for $\omega \in \Omega$ under the assumption of a Laplace model.
Solution:

Since $|\Omega| = 3^3 = 27$, we have $p(\omega) = 1/27$ for all $\omega \in \Omega$.

(c) [**2 Points**]
Compute the probability that all three people go to the same floor.
Solution:

Since there are three floors, the cardinality of this event E_1 is three. Thus, its probability is $\mathbb{P}(E_1) = 3/27 = 1/9$.

(d) [**2 Points**]
Compute the probability that they go to exactly two different floors.
Solution:

There are $\binom{3}{2} \times 2 \times \binom{3}{2}$ possible configurations in this event E_2. Hence, $\mathbb{P}(E_2) = 3^2 \times 2/3^3 = 2/3$.

(e) [**2 Points**]
Compute the probability that they all go to different floors.
Solution:

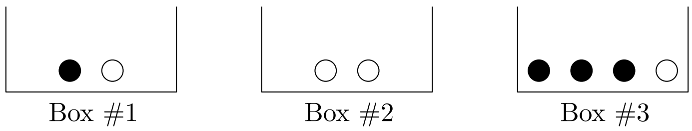
There are 3! possibilities for this event E_3. Hence, the probability is $\mathbb{P}(E_3) = 3!/3^3 = 2/9$.

Alternative: $\Omega = E_1 \cup E_2 \cup E_3$ is a disjoint union. Thus, $\mathbb{P}(E_1) = 1 - \mathbb{P}(E_2) - \mathbb{P}(E_3) = 2/9$.

3. [**7 Points**]

Consider three boxes such that

- Box #1 contains one black and one white ball;
- Box #2 contains two white balls;
- Box #1 contains three black and one white ball.



We first choose randomly a box and then a ball from this box. From a chosen box, the balls have the same probability to be drawn. Let $W = \{\text{The ball drawn is white}\}$.

(a) [**2 Points**]

We assume in this question that the boxes have the same probability to be chosen. Compute the probability of the event W.

Solution:

We have $\mathbb{P}(W) = \sum_{i=1}^{3} \mathbb{P}(W|\text{Box \#i is chosen}) \times \frac{1}{3} = \frac{1}{3}\left(\frac{1}{2} + 1 + \frac{1}{4}\right) = \frac{7}{12}$.

(b) [**2 Points**]

We assume now that $\mathbb{P}(\{\text{Picking Box \#1}\}) = \mathbb{P}(\{\text{Picking Box \#2}\}) = 1/4$. Compute the probability that the ball drawn is black.

Solution:

Let $B = \{\text{The ball drawn is black}\}$. Then,

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B|\text{Box \#1}) \times \frac{1}{4} + \mathbb{P}(B|\text{Box \#2}) \times \frac{1}{4} + \mathbb{P}(B|\text{Box \#3}) \times \frac{1}{2} \\ &= \frac{1}{2}\frac{1}{4} + 0\frac{1}{4} + \frac{3}{4}\frac{1}{2} = \frac{1}{2}.\end{aligned}$$

(c) [**3 Points**]

We assume again that the boxes have the same probabilities of being chosen. Given that the ball drawn is white, what is the conditional probability that it was drawn from Box #2?

Solution:

We have

$$\begin{aligned}\mathbb{P}(\text{Box }\#2|W) &= \frac{\mathbb{P}(W|\text{Box }\#2)\mathbb{P}(\text{Box }\#2)}{\mathbb{P}(W)} \\ &= \frac{\mathbb{P}(W|\text{Box }\#2)\mathbb{P}(\text{Box }\#2)}{\sum_{i=1}^{3}\mathbb{P}(W|\text{Box }\#\text{i})\mathbb{P}(\text{Box }\#\text{i})} = \frac{1 \times \frac{1}{3}}{\frac{1}{3}\left(\frac{1}{2}+1+\frac{1}{4}\right)} = \frac{4}{7}.\end{aligned}$$

4. [6 Points]

Consider a square with a random length X. We assume that $X \sim \mathcal{U}([0,a])$ is distributed uniformly for some $a > 0$. Let A denote the area of the square.

(a) **[2 Points]**
Compute the expected value $\mathbb{E}[A]$ and the variance $\mathbb{V}(A)$.
Solution:

We have
$$\mathbb{E}[A] = \mathbb{E}[X^2] = \int_0^a \frac{x^2}{a} dx = \frac{a^2}{3}$$
and $\mathbb{V}(A) = \mathbb{E}[A^2] - \mathbb{E}[A]^2$ with
$$\mathbb{E}[A^2] = \mathbb{E}[X^4] = \int_0^a \frac{x^4}{a} dx = \frac{a^4}{5},$$
so $\mathbb{V}(A) = \frac{a^4}{5} - \frac{a^4}{9} = \frac{4a^4}{45}$.

(b) **[2 Points]**
Compute the cumulative distribution function of A.
Solution:

We have
$$F_A(x) = \mathbb{P}(X^2 \leq x) = \left\{\begin{matrix} 0 & \text{if } x < 0 \\ \mathbb{P}(X \leq \sqrt{x}) & \text{if } x \geq 0 \end{matrix}\right\} = \begin{cases} 0 & \text{if } x < 0 \\ \sqrt{x}/a & \text{if } 0 \leq x < a^2 \\ 1 & \text{if } x \geq a^2. \end{cases}$$

(c) **[2 Points]**
Determine all $a > 0$ such that $\mathbb{P}(A > 1) \geq 1/2$.
Solution:

We have
$$\mathbb{P}(A > 1) = 1 - F_A(1) = \begin{cases} 0 & \text{if } 1 \geq a^2 \\ 1 - 1/a & \text{if } 1 < a^2. \end{cases}$$
Thus, $\mathbb{P}(A > 1) \geq 1/2$ iff $1 - 1/a \geq 1/2$ and $a > 1$ iff $a \geq 2$.

5. [8 Points]

Let $X_1, \dots, X_n$ be i.i.d. $\sim \text{Pois}(\lambda)$ for some $\lambda > 0$.

(a) **[1 Point]**

Recall the definition of convergence in probability.

Solution:

We say that a sequence of random variables $(X_n)_{n\geq 1}$ converges in probability to a random variable X if $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ for all $\varepsilon > 0$.

(b) **[2 Points]**

By using Chebyshev's inequality, show that $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \lambda$ as $n \to \infty$.

Solution:

Let $\varepsilon > 0$. Then,

$$\mathbb{P}(|\bar{X}_n - \lambda| > \varepsilon) \leq \frac{\mathbb{V}(\bar{X}_n)}{\varepsilon^2} = \frac{\lambda}{n\varepsilon^2} \to 0 \qquad \text{as } n \to \infty.$$

(c) **[2 Points]**

Write down the Central Limit Theorem for the random variable $\bar{X}_n$.

Solution:

Since $\mathbb{E}[X_i] = \lambda < \infty$ and $\mathbb{V}(X_i) = \lambda < \infty$, the CLT applies and we have

$$\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{as } n \to \infty.$$

(d) **[3 Points]**

Suppose that some hotel opens only for 100 days in any given year. For $i \in \{1, \dots, 100\}$, let X_i = The number of people the hotel receives on day $\#i$. Assuming that $X_1, \dots, X_{100}$ are i.i.d. $\sim \text{Pois}(9)$, give an approximation of

$$\mathbb{P}\left(840 < \sum_{i=1}^{100} X_i \leq 960\right).$$

You may use: $\Phi(1) \approx 0.84$, $\Phi(3/2) \approx 0.93$, and $\Phi(2) \approx 0.97$, where $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-x^2/2}dx$ for $z \in \mathbb{R}$.

Solution:

Note that

$$840 < \sum_{i=1}^{100} X_i \leq 960 \qquad \Longleftrightarrow \qquad -2 < \frac{\sum_{i=1}^{100} X_i - 100 \times 9}{10 \times 3} \leq 2.$$

Hence, by the CLT,

$$\mathbb{P}\left(840 < \sum_{i=1}^{100} X_i \leq 960\right) \approx \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.94.$$

Part II: Statistics

6. [**9 Points**]

Consider the parametric model $\mathcal{P} = \{P_\theta \mid \theta \in (0,\infty)\}$, where P_θ admits the density

$$f_\theta(x) = \theta(1-x)^{\theta-1}\mathbb{1}_{x\in[0,1]}.$$

(a) [**2 Points**]
Compute $\mathbb{E}_\theta[X]$, where $X \sim P_\theta$.
Solution:

We have

$$\mathbb{E}_\theta[X] = \theta\int_0^1 x(1-x)^{\theta-1}dx = \theta\int_0^1 (x-1+1)(1-x)^{\theta-1}dx$$
$$= -\theta\int_0^1 (1-x)^{\theta}dx + \theta\int_0^1 (1-x)^{\theta-1}dx = \frac{1}{\theta+1}.$$

Alternative: f_θ is the density of a Beta distribution with parameters $\alpha = 1$ and $\beta = \theta$. In particular, $\mathbb{E}_\theta[X] = \alpha/(\alpha+\beta) = 1/(1+\theta)$.

(b) [**2 Points**]
Construct the moment estimator of θ_0 based on i.i.d. $X_1, \ldots, X_n \sim P_{\theta_0}$.
Solution:

Let $\hat{\theta}_n$ be the moment estimator. Then, $1/(\hat{\theta}_n + 1) = \bar{X}_n$, so $\hat{\theta}_n = 1/\bar{X}_n - 1$.

(c) [**2 Points**]
Using the appropriate theorems, show that the moment estimator obtained in question (b) converges in probability to θ_0.
Solution:

By the WLLN, $\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}_{\theta_0}[X] = 1/(\theta_0+1)$. Put $g(x) = 1/x - 1$ for $x \in (0,\infty)$. This function is continuous on $(0,\infty)$. By the continuous mapping theorem, we have $g(\bar{X}_n) \xrightarrow{\mathbb{P}} g(1/(\theta_0+1)) = \theta_0$, that is, $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

(d) [**3 Points**]
For $\mathbb{X} = (X_1, \ldots, X_n)$ and $X_1, \ldots, X_n$ i.i.d. $\sim P_{\theta_0}$, write down the log-likelihood $l_{\mathbb{X}}(\theta)$ for $\theta \in (0,\infty)$ and find the MLE. (You can assume that $X_i \in (0,1)$ for all $i \in \{1, \ldots, n\}$.)
Solution:

We have $L_{\mathbb{X}}(\theta) = \prod_{i=1}^n f_\theta(X_i)$ and

$$l_{\mathbb{X}}(\theta) = \sum_{i=1}^n \log(\theta(1-X_i)^{\theta-1}) = n\log(\theta) + (\theta-1)\sum_{i=1}^n \log(1-X_i).$$

Thus,

$$0 = l'_{\mathbb{X}}(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log(1 - X_i) \qquad \Longleftrightarrow \qquad \theta = -\frac{n}{\sum_{i=1}^{n} \log(1 - X_i)} =: \tilde{\theta}_n.$$

Note that $l_{\mathbb{X}}$ is strictly concave and, hence, the critical point $\tilde{\theta}_n$ has to be the MLE.

7. [6 Points]

Consider the uniform distribution $\mathcal{U}([\theta, \theta+1])$ for $\theta \in \mathbb{R}$.

(a) **[1 Point]**
Compute the expected value $\mathbb{E}_\theta[X]$ for $X \sim \mathcal{U}([\theta, \theta+1])$ (do not use the sheet of formulas).
Solution:

We have $\mathbb{E}_\theta[X] = \int_\theta^{\theta+1} x dx = ((\theta+1)^2 - \theta^2)/2 = \theta + 1/2.$

(b) **[1 Point]**
Compute the variance $\mathbb{V}_\theta(X)$ for $X \sim \mathcal{U}([\theta, \theta+1])$ (do not use the sheet of formulas).
Solution:

Note that $X = Y + \theta$ with $Y \sim \mathcal{U}(0,1)$. Hence,

$$\mathbb{V}_\theta(X) = \mathbb{V}_\theta(Y) = \mathbb{V}(Y) = \int_0^1 y^2 dy - \left(\int_0^1 y dy\right)^2 = 1/12.$$

(c) **[2 Points]**
Let $X_1, \ldots, X_n$ be i.i.d. $\sim \mathcal{U}([\theta_0, \theta_0+1])$. State the Central Limit Theorem for $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ in this case.
Solution:

The expected value and the variance are finite. Hence, by the CLT, $\frac{\sqrt{n}(\bar{X}_n - (\theta_0 + 1/2))}{\sqrt{1/12}} \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$.

(d) **[2 Points]**
Based on question (c), construct a bilateral and symmetric confidence interval for θ_0 with asymptotic level $1-\alpha$ for $\alpha \in (0,1)$.
Solution:

We have

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\sqrt{n}(\bar{X}_n - (\theta_0 + 1/2))}{\sqrt{1/12}} \leq z_{1-\alpha/2}\right) \approx 1-\alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$. Thus,

$$\mathbb{P}\left(\bar{X}_n - \frac{1}{2} - \frac{z_{1-\alpha/2}}{\sqrt{12n}} \leq \theta_0 \leq \bar{X}_n - \frac{1}{2} + \frac{z_{1-\alpha/2}}{\sqrt{12n}}\right) \approx 1-\alpha,$$

and the confidence interval is $\left[\bar{X}_n - \frac{1}{2} - \frac{z_{1-\alpha/2}}{\sqrt{12n}}, \bar{X}_n - \frac{1}{2} + \frac{z_{1-\alpha/2}}{\sqrt{12n}}\right]$.

8. [8 Points]

Consider the probability density function

$$f_\theta(x) = \frac{c}{x^3}\mathbb{1}_{\{x\geq\theta\}}$$

with $\theta \in (0,\infty)$ and $c > 0$.

(a) **[1 Points]**

Determine $c > 0$ as a function of θ.

Solution:

We have $\int_0^\infty cx^{-3}dx = 1$ iff $c = 2\theta^2$.

(b) **[2 Points]**

Let $\mathbb{X} = (X_1,\dots,X_n)$, where $X_1,\dots,X_n$ are i.i.d. $\sim f_{\theta_0}$. Write down the likelihood function $L_{\mathbb{X}}(\theta)$ and show that the MLE is $\min_{1\leq i\leq n} X_i$.

Solution:

We have

$$L_{\mathbb{X}}(\theta) = \prod_{i=1}^n \frac{2\theta^2}{X_i^3}\mathbb{1}_{\{X_i\geq\theta\}} = \frac{2^n\theta^{2n}}{\prod_{i=1}^n X_i^3}\mathbb{1}_{\{\min_{1\leq i\leq n} X_i\geq\theta\}}. \tag{1}$$

Thus, the MLE is indeed equal to $\min_{1\leq i\leq n} X_i$.

(c) **[3 Points]**

Compute the cumulative distribution function of $\min_{1\leq i\leq n} X_i$, where, as before, $X_1,\dots,X_n$ are i.i.d. $\sim f_{\theta_0}$.

Solution:

We have

$$\begin{aligned}\mathbb{P}\left(\min_{1\leq i\leq n} X_i \leq t\right) &= 1 - \mathbb{P}\left(\min_{1\leq i\leq n} X_i > t\right)\\ &= 1 - \mathbb{P}(X_1 > t,\dots,X_n > t) = 1 - \mathbb{P}(X_1 > t)^n = 1-(1-F(t))^n,\end{aligned}$$

where

$$F(t) = \mathbb{P}(X_1 \leq t) = \begin{Bmatrix} 0 & \text{if } t < \theta_0 \\ \int_{\theta_0}^t 2\theta_0^2 x^{-3}dx & \text{if } t \geq \theta_0 \end{Bmatrix} = \left(1 - \left(\frac{\theta_0}{t}\right)^2\right)\mathbb{1}_{\{t\geq\theta_0\}}.$$

We conclude that

$$\mathbb{P}\left(\min_{1\leq i\leq n} X_i \leq t\right) = \left(1 - \left(\frac{\theta_0}{t}\right)^{2n}\right)\mathbb{1}_{\{t\geq\theta_0\}}.$$

(d) **[2 Points]**

For a fixed $\varepsilon > 0$, compute $\mathbb{P}(|\min_{1\le i\le n} X_i - \theta_0| > \varepsilon)$ and deduce that the MLE converges to θ_0 in probability as $n \to \infty$.

Solution:

We have

$$\mathbb{P}\left(\left|\min_{1\le i\le n} X_i - \theta_0\right| > \varepsilon\right) = \mathbb{P}\left(\min_{1\le i\le n} X_i > \theta_0 + \varepsilon\right) = 1 - \mathbb{P}\left(\min_{1\le i\le n} X_i \le \theta_0 + \varepsilon\right) = \left(\frac{\theta_0}{\theta_0 + \varepsilon}\right)^{2n}.$$

Thus, $\lim_{n\to\infty} \mathbb{P}(|\min_{1\le i\le n} X_i - \theta_0| > \varepsilon) = 0$, that is, $\min_{1\le i\le n} X_i \xrightarrow{\mathbb{P}} \theta_0$ as $n \to \infty$.

9. [9 Points]

Consider $X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(\theta, \sigma_0^2)$, where $\sigma_0 > 0$ is known and $\theta \in \mathbb{R}$. We want to test $H_0\colon \theta = 0$ versus $H_1\colon \theta = 1$. $(\star)$

(a) **[2 Points]**
Recall the Neyman-Pearson test of level α for testing a simple null hypothesis $H_0\colon p = p_0$ versus a simple alternative hypothesis $H_1\colon p = p_1$, where p is the density of an observed sample with respect to some σ-finite dominating measure.

Solution:

Let X be the observed sample. The NP-test of level α is

$$\Phi_{NP}(X) = \begin{cases} 1 & \text{if } \frac{p_1(X)}{p_0(X)} > k_\alpha \\ q_\alpha & \text{if } \frac{p_1(X)}{p_0(X)} = k_\alpha \\ 0 & \text{if } \frac{p_1(X)}{p_0(X)} < k_\alpha, \end{cases}$$

where k_α is the $(1-\alpha)$-quantile of $p_1(X)/p_0(X)$ under H_0 and q_α is such that

$$\mathbb{E}_{p_0}[\Phi_{NP}(X)] = \mathbb{P}_{p_0}\left(\frac{p_1(X)}{p_0(X)} > k_\alpha\right) + q_\alpha \mathbb{P}_{p_0}\left(\frac{p_1(X)}{p_0(X)} = k_\alpha\right) = \alpha.$$

(b) **[3 Points]**
Find the NP-test of level α for the testing problem $(\star)$.

Solution:

We have

$$\begin{aligned} \frac{p_1(X_1,\dots,X_n)}{p_0(X_1,\dots,X_n)} &= \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(X_i-1)^2}{2\sigma_0^2}\right)}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{X_i^2}{2\sigma_0^2}\right)} \\ &= \exp\left(-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n (X_i-1)^2 - X_i^2\right)\right) \\ &= \exp\left(\frac{1}{\sigma_0^2}\sum_{i=1}^n X_i\right)\exp\left(-\frac{n}{2\sigma_0^2}\right). \end{aligned}$$

Thus,

$$\Phi_{NP}(X) = \begin{cases} 1 & \text{if } \frac{1}{\sigma_0^2}\sum_{i=1}^n X_i > \tilde{k}_\alpha \\ \tilde{q}_\alpha & \text{if } \frac{1}{\sigma_0^2}\sum_{i=1}^n X_i = \tilde{k}_\alpha \\ 0 & \text{if } \frac{1}{\sigma_0^2}\sum_{i=1}^n X_i > \tilde{k}_\alpha. \end{cases}$$

Here, $\tilde{q}_\alpha$ can be taken to be zero because the distribution is continuous under H_0 (actually, it is always continuous since it is Gaussian). The NP-test can be rewritten as

$$\Phi_{NP}(X) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}\bar{X}_n}{\sigma_0} > t_\alpha \\ 0 & \text{otherwise} \end{cases}$$

with t_α such that $\mathbb{P}_{\theta=0}(\sqrt{n}\bar{X}_n/\sigma_0 > t_\alpha) = \alpha$, that is, t_α is the $(1-\alpha)$-quantile of $\mathcal{N}(0,1)$.

(c) **[2 Points]**
Give the power of the NP-test obtained in question (b) and show that it converges to 1 as $n \to \infty$.

Solution:

The power is

$$\beta_n = \mathbb{P}_{\theta=1}\left(\frac{\sqrt{n}\bar{X}_n}{\sigma_0} > t_\alpha\right) = \mathbb{P}_{\theta=1}\left(\frac{\sqrt{n}(\bar{X}_n - 1)}{\sigma_0} > t_\alpha - \frac{\sqrt{n}}{\sigma_0}\right) = \mathbb{P}_{\theta=1}\left(Z > t_\alpha - \frac{\sqrt{n}}{\sigma_0}\right)$$

with $Z \sim \mathcal{N}(0,1)$. Thus,

$$\lim_{n\to\infty} \beta_n = 1 - \lim_{n\to\infty} \mathbb{P}_{\theta=1}\left(Z \le t_\alpha - \frac{\sqrt{n}}{\sigma_0}\right) = 1.$$

(d) **[2 Points]**
Argue that the NP-test obtained in question (b) is UMP for testing $H_0\colon \theta = 0$ versus $H_1\colon \theta > 0$.

Solution:

The test Φ_{NP} does not involve the particular value 1 for θ in H_1. Since Φ_{NP} is UMP for $H_0\colon \theta = 0$ versus any alternative $H_1\colon \theta = \theta_1$ with $\theta_1 > 0$, we conclude that its power $\beta(\theta_1)$ has to be maximal. This means that it has to be UMP for testing $H_0\colon \theta = 0$ versus $H_1\colon \theta > 0$.

10. [6 Points]

100 students who attended the lectures on Probability and Statistics were asked the following questions:

- Q1: Did you like the lectures? (yes/no)
- Q2: What was your preferred mode of attendance? (presence/online)

Put

$$X = \begin{cases} 1 & \text{if the answer to Q1 is 'yes'} \\ 2 & \text{if the answer to Q1 is 'no'} \end{cases} \qquad Y = \begin{cases} 1 & \text{if the answer to Q2 is 'presence'} \\ 2 & \text{if the answer to Q2 is 'online'} \end{cases}$$

The goal is to test whether there is an association between X and Y.

(a) **[1 Points]**
Describe the testing problem mathematically.
Solution:

We want to test H_0: X and Y are independent versus H_1: They are not independent.

(b) **[3 Points]**
Write down the appropriate test-statistic and the related test of asymptotic level α.
Solution:

The test-statistic is

$$Q_n = n\frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1+}N_{2+}N_{+1}N_{+2}}$$

with $N_{ij} = \#\{k \mid (X_k, Y_k) = (i.j)\}$ and $N_{i+} = N_{i1} + N_{i2}$, $N_{+j} = N_{1j} + N_{2j}$ for the data given by i.i.d. copies (X_k, Y_k) of (X, Y). The test is

$$\Phi(N_{11}, N_{12}, N_{21}, N_{22}) = \begin{cases} 1 & \text{if } Q_n > q_{1-\alpha} \\ 0 & \text{otherwise} \end{cases}$$

with $q_{1-\alpha}$ the $(1-\alpha)$-quantile of $\chi^2_{(1)}$.

(c) **[2 Points]**
Take $\alpha = 0.05$. What is the decision you make using the test from question (b) if the survey results yielded the following contingency table?

	Y	
X	40	30
	10	20

You may use:

- the 0.95-quantile of $\chi^2_{(1)} \approx 3.84$,
- the 0.975-quantile of $\chi^2_{(1)} \approx 5.02$,

- the 0.95-quantile of $\chi^2_{(2)} \approx 5.99$,
- the 0.975-quantile of $\chi^2_{(2)} \approx 7.37$.

Solution:

We have

$$Q_n = 100 \times \frac{(40 \times 20 - 30 \times 10)^2}{50 \times 50 \times 70 \times 30} = \frac{100}{21} \approx 4.76 > 3.84,$$

so we reject H_0.