Q1.) a.) $X_i \sim \text{Bernoulli}(p), p \in (0,1)$; $Y_i \sim \text{Bernoulli}(q), q \in (0,1)$ - known, $X_1, \dots, X_n, Y_1, \dots, Y_n$ - mutually independent

$$P_{Y,Z}(y,z \mid p,q) = \begin{cases} q^y (1-q)^{1-y} \{ p^z (1-p)^{1-z} \}^y & , (y,z) \in \{(0,0),(1,0),(1,1)\} \\ 0 & , \text{otherwise} \end{cases}$$

The likelihood is

$$L(p,q \mid \underline{y}, \underline{z}) = \prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i} \{ p^{z_i}(1-p)^{1-z_i} \}^{y_i} = q^{\sum_{i=1}^n y_i} (1-q)^{\sum_{i=1}^n (1-y_i)} p^{\sum_{i=1}^n z_i \sum_{i=1}^n y_i} (1-p)^{\sum_{i=1}^n (1-z_i) \sum_{i=1}^n y_i},$$

known; observed random samples of size n

The log-likelihood is

$$\ell(p,q \mid \underline{y}, \underline{z}) = \log L(p,q \mid \underline{y}, \underline{z}) = \log q \cdot \sum_{i=1}^n y_i + \log(1-q) \cdot \sum_{i=1}^n (1-y_i) + \log p \cdot \sum_{i=1}^n z_i \sum_{i=1}^n y_i + \log(1-p) \cdot \sum_{i=1}^n (1-z_i) \sum_{i=1}^n y_i$$

known

The aim is to maximize the likelihood function, or in our case the log-likelihood function. $\log(x)$ is a strictly increasing function of x, therefore the maximising argument of $\ell(p,q \mid \underline{y}, \underline{z})$ and $L(p,q \mid \underline{y}, \underline{z})$ will coincide exactly.

Solving the following equation will give us a stationary point:

$$\frac{\partial \ell(p,q \mid \underline{y}, \underline{z})}{\partial p} = 0 \quad \Rightarrow \quad \frac{\sum_{i=1}^n z_i \sum_{i=1}^n y_i}{p} - \frac{\sum_{i=1}^n (1-z_i) \sum_{i=1}^n y_i}{1-p} = 0, \; p \in (0,1)$$

$$(1-p) \sum_{i=1}^n z_i \sum_{i=1}^n y_i - p \sum_{i=1}^n (1-z_i) \sum_{i=1}^n y_i = 0 \Rightarrow p \left[\sum_{i=1}^n z_i \sum_{i=1}^n y_i + \sum_{i=1}^n (1-z_i) \sum_{i=1}^n y_i \right] = \sum_{i=1}^n z_i \sum_{i=1}^n y_i$$

Hence

$$p = \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n z_i + \sum_{i=1}^n (1-z_i)} = \frac{1}{n} \cdot \sum_{i=1}^n z_i$$

If we take the second derivative of the log-likelihood function we get:

$$\frac{d^2 \ell(p,q \mid \underline{y}, \underline{z})}{dp^2} = - \frac{\sum_{i=1}^n z_i \sum_{i=1}^n y_i}{p^2} - \frac{\sum_{i=1}^n (1-z_i) \sum_{i=1}^n (y_i)}{(1-p)^2} < 0, \; \forall p \in (0,1), \; y_i \in \{0,1\}, z_i \in \{0,1\}$$

Since ℓ is continuous on $0 \le p \le 1$ and $\ell''(p) < 0$ for $0 < p < 1$, $\ell(p)$ is strictly concave

$\Rightarrow$ any local maximum that we find will be the unique global maximum

$\Rightarrow$ the MLE is $\hat{p} = p = \frac{1}{n} \cdot \sum_{i=1}^n z_i$

b.) $\tilde{p} = \frac{1}{nq} \sum_{i=1}^{n} z_i$

$\tilde{p}$ would be unbiased if we can show $E\tilde{p} = p$.

$E\tilde{p} = E\left[\frac{1}{nq} \sum_{i=1}^{n} z_i\right] = \frac{1}{nq} \cdot E \sum_{i=1}^{n} z_i = \frac{1}{nq} \cdot \sum_{i=1}^{n} E z_i = \frac{1}{nq} \cdot \sum_{i=1}^{n} E[X_i \cdot Y_i]$

By problem statement $X_1, \dots, X_n, Y_1, \dots, Y_n$ are mutually independent $\Rightarrow E[X_i Y_i] = EX_i \cdot EY_i$

$\Rightarrow E\tilde{p} = \frac{1}{nq} \cdot \sum_{i=1}^{n} EX_i EY_i$

By problem statement $X_i \sim \text{Bernoulli}(p) \Rightarrow EX_i = p$ and $Y_i \sim \text{Bernoulli}(q) \Rightarrow EY_i = q$

$\Rightarrow E\tilde{p} = \frac{1}{nq} \cdot \sum_{i=1}^{n} p \cdot q = \frac{1}{nq} \cdot n \cdot p \cdot q = p \quad \Rightarrow \tilde{p}$ is unbiased

$Var[\tilde{p}] = Var\left[\frac{1}{nq} \sum_{i=1}^{n} z_i\right] = \frac{1}{n^2q^2} \cdot Var\left[\sum_{i=1}^{n} z_i\right] = \frac{1}{n^2q^2} \cdot E\left[\sum_{i=1}^{n} z_i - E\sum_{i=1}^{n} z_i\right]^2 =$

$= \frac{1}{n^2q^2} E\left[\sum_{i=1}^{n} (z_i - Ez_i)\right]^2 = \frac{1}{n^2q^2} E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} (z_i - Ez_i)(z_j - Ez_j)\right] =$

$= \frac{1}{n^2q^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[(z_i - Ez_i)(z_j - Ez_j)\right]$

We have n terms s.t.: $E(z_i - Ez_i)^2 = Var z_i \quad (i=j)$

For the remaining terms $(i \neq j)$ we have:

$E[(z_i - Ez_i) \cdot (z_j - Ez_j)] = Cov(z_i, z_j) = 0$ because $X_1, \dots, X_n, Y_1, \dots, Y_n$ are mutually independent which leads to $z_i = X_i \cdot Y_i$, $i = 1, \dots, n$ to also be mutually independent.

$$\Rightarrow Var[\tilde{p}] = \frac{1}{n^2q^2} \cdot n \cdot Var[z_i] = \frac{1}{n \cdot q^2} \cdot Var z_i$$

$Var[z_i] = Var[X_i \cdot Y_i] = E[X_i \cdot Y_i]^2 - (E[X_i \cdot Y_i])^2 = E[X_i^2 \cdot Y_i^2] - (E[X_i \cdot Y_i])^2 =$

$= EX_i^2 \cdot EY_i^2 - (EX_i)^2 (EY_i)^2 = [Var X_i + (EX_i)^2][Var Y_i + (EY_i)^2] - (EX_i)^2 (EY_i)^2$

$X_i \sim \text{Bernoulli}(p) \Rightarrow EX_i = p$, $Var X_i = p(1-p)$

$Y_i \sim \text{Bernoulli}(q) \Rightarrow EY_i = q$, $Var Y_i = q(1-q)$

$\Rightarrow Var z_i = [p(1-p) + p^2][q(1-q) + q^2] - p^2 \cdot q^2 = p \cdot q - p^2q^2 = pq(1-pq)$

$$\Rightarrow Var[\tilde{p}] = \frac{1}{n \cdot q^2} \cdot p \cdot q \cdot (1 - p \cdot q) = \frac{1}{nq} \cdot p(1-pq)$$

c) Let p_0 = true value of p, $\hat{p}$ = MLE

For sufficiently large n (given by problem statement) and "regularity conditions"

$$\Rightarrow \hat{p} \underset{\text{approximately distributed}}{\overset{a}{\sim}} \mathcal{N}(p_0, \underbrace{I_n(p_0)}^{-1}) \qquad \text{i.i.d. } I_n(p) = n \cdot I(p)$$

$I_n(p_0)$ – Fisher information

where $I_n(p) := E_p\left[\underbrace{-\frac{d^2}{dp^2}}\ell(p; y, z)\right]$

- measures how curved the log-likelihood is
- the bigger curvature the more information ℓ contains and the more precise the MLE

$$I_n(p_0)^{0.5}(\hat{p} - p_0) \xrightarrow[n\to\infty]{} \mathcal{N}(0,1)$$ – asymptotic sampling distribution of the MLE

From a.) we have
$$\frac{d^2\ell(p,q|y,z)}{dp^2} = -\frac{\sum_{i=1}^n z_i \sum_{i=1}^n y_i}{p^2} - \frac{\sum_{i=1}^n (1-z_i)\sum_{i=1}^n (y_i)}{(1-p)^2} = \sum_{i=1}^n \left[-\frac{z_i y_i}{p^2} - \frac{(1-z_i)y_i}{(1-p)^2}\right]$$

$\Rightarrow$ The Fisher information for all observations is

$$I_n(p) = E\left[-\frac{d^2}{dp^2}\ell(p)\right] = \sum_{i=1}^n \left[\frac{E[z_i y_i]}{p^2} + \frac{E[(1-z_i)y_i]}{(1-p)^2}\right] =$$

// $EY_i^2 = Var Y_i + (EY_i)^2$

$$= \sum_{i=1}^n \left[\frac{EX_i \cdot EY_i^2}{p^2} + \frac{EY_i}{(1-p)^2} - \frac{EX_i \cdot EY_i^2}{(1-p)^2}\right] =$$

$$= \sum_{i=1}^n \left[\frac{p[q(1-q)+q^2]}{p^2} + \frac{q}{(1-p)^2} - \frac{p[q(1-q)+q^2]}{(1-p)^2}\right] =$$

$$= \sum_{i=1}^n \left[\frac{p \cdot q}{p^2} + \frac{q}{(1-p)^2} - \frac{pq}{(1-p)^2}\right] = \sum_{i=1}^n \left[\frac{q(1-2p)}{p(1-p)^2} + \frac{q}{(1-p)}\right] = \sum_{i=1}^n \frac{q}{p(1-p)} = \frac{nq}{p(1-p)}$$

$\Rightarrow$ for large n, $\hat{p}$ is approximately $\mathcal{N}\left(p, \frac{p(1-p)}{nq}\right)$

$\tilde{p} = \frac{1}{nq}\sum_{i=1}^n z_i = \frac{\bar{z}}{q}$, $E\tilde{p} = p$, $Var\,\tilde{p} = \frac{p(1-pq)}{nq}$ $\qquad$ $Var\,z = pq(1-pq)$, $Ez = pq$

CLT: $\bar{z} \sim \mathcal{N}\left(Ez, \frac{Var\,z}{n}\right)$ or if $g(\bar{z}) = \tilde{p} = \frac{1}{nq}\sum_{i=1}^n z_i = \frac{\bar{z}}{q} \Rightarrow g'(z) = \frac{1}{q}$

$$\tilde{p} = g(\bar{z}) \sim \mathcal{N}\left(g(Ez), \frac{Var(z) \cdot g'(Ez)^2}{n}\right) =$$

$$= \mathcal{N}\left(g(pq), \frac{pq(1-pq)}{n} \cdot \frac{1}{q^2}\right) = \mathcal{N}\left(p, \frac{p(1-pq)}{nq}\right)$$

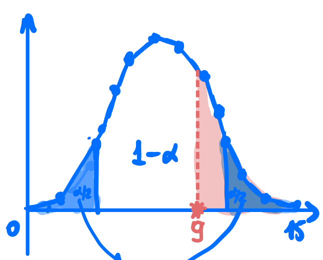
$\frac{\overbrace{p(1-p)}^{>0}}{nq} < \frac{\overbrace{p(1-pq)}^{>0}}{nq} \Rightarrow$ the variance (and standard deviation) is bigger for $\tilde{p}$ and smaller for $\hat{p}$

In terms of Fisher information: the Fisher information is the inverse of the variance, which would mean the bigger the variance the less the Fisher information. If we state that the bigger the curvature the better, this leads to stating the bigger the Fisher information the better. This would mean to chose $\hat{p}$ which also aligns with the basic logic to chose the estimator with asymptotic distribution having the smaller variance, and hence the smaller standard deviation.

Q2.) a.) Let the test statistic be the number of students that pass the exam, X. The sampling distribution is a Binomial distribution $B(n,p)$, or $X \sim B(n,p)$, where n is the number of students taking the same exam, and p is the probability for a student to pass the exam.

First, let's assume that H_0 is true $\Rightarrow X \sim B(15, 0.5)$.

We are performing a two-tailed test, so for the critical region we have:



Critical region

```python
from scipy.stats import binom

n = 15
p = 0.5
r_values = list(range(n + 1))

dist = [binom.cdf(r, n, p) for r in r_values]

print("r\tp(r)")
for r_value, dist_value in zip(r_values, dist):
    print(f"{r_value}\t{dist_value:.4f}")
```

this is a two-tailed test, that's why we use $\alpha/2$, only referring to the upper tail

We will test the upper tail (the expectation is $15 \cdot \frac{1}{2} < 9$, which means that the upper tail will have lower probability).

$P_x(X \geq 9) = 1 - P_x(X < 9) = 1 - P_x(X \leq 8) = 1 - 0.6964 = 0.3036 > \frac{\alpha}{2} = 0.025$, so the observed statistic doesn't lie within the critical region $\Rightarrow$ the result is not significant $\Rightarrow$ Fail to reject the null hypothesis H_0.

There is insufficient evidence to suggest that the probability of passing the exam is different to 0.5.

b.) This time we have $n=500$ and we are considering the same hypothesis:

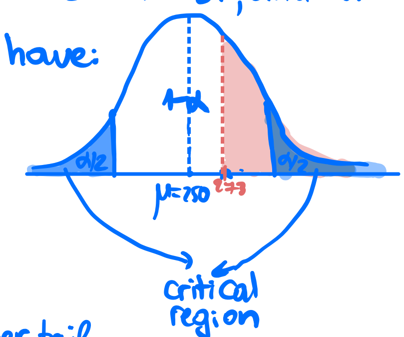
$$\begin{cases} H_0: p=0.5 \\ H_1: p\neq 0.5 \end{cases}$$

. Let the test statistic be the number of students that pass the exam, X.

If we assume H_0 is true we have that while p stayed the same compared to a.) the number of students has increased. From the Central Limit Theorem the Binomial distribution with n trials and probability of success p, will get closer to normal distribution as n increases, and p is fixed. $B(n,p)\xrightarrow[n\to\infty]{} \mathcal{N}(np, np(1-p))$

$\Rightarrow$ if H_0 is true, $X\sim\mathcal{N}(500\cdot\frac{1}{2}, 500\cdot\frac{1}{2}\cdot\frac{1}{2})=\mathcal{N}(250,125)$, $\mu=250, \sigma^2=125$

Again, we are doing a two-sided test, and for the critical region we have:



Again, we'll test the upper tail.

Rewriting the problem using the continuity correction factor (for the approximation to $\mathcal{N}$)

CCF

We need to find $P_X(X\geq 273)$ but after using the CCF it becomes $P_X(X\geq 272.5)$

corrected value

Finding the z-score: $z=\frac{X-\mu}{\sigma}=\frac{272.5-250}{\sqrt{125}}=\frac{22.5}{15}=1.5$

Using a z table $\Rightarrow P(Z\geq 1.5)=1-0.9332=0.0668>\frac{\alpha}{2}=0.025$ (analogically to a)) $\Rightarrow$

the observed statistic doesn't lie within the critical region $\Rightarrow$

the result is not significant $\Rightarrow$ Fail to reject the null hypothesis H_0.

There is insufficient evidence to suggest that the probability of passing the exam is different to 0.5.