Final assignment

Prediction

We need to calculate P(survived | 2 positive nodes) =

$$\frac{P(2\text{ positive nodes}|\text{ survived})\mathrm{P}(\text{survived})}{P(2\text{ positive nodes}|\text{ survived})\mathrm{P}(\text{survived}) + P(2\text{ positive nodes}|\text{ died})\mathrm{P}(\text{died})}$$

<u>For the died distribution</u>:

$$\mathrm{P}(\text{died}) = \frac{81}{306} = 0.265$$

And according to geometric distribution 2nd version (with 0s):

Calculate p:

$$\frac{1-p}{p} = 7.5 \Rightarrow 7.5p = 1-p \Rightarrow 8.5p = 1 \Rightarrow p = 0.118$$

calculate $P(2\text{ positive nodes}|\ died)$:

$$PMF(x=2) = (1-p)^2 p = (1-0.118)^2 \cdot 0.118 = 0.092$$

<u>For the survived distribution</u>:

$$\mathrm{P}(\text{survived}) = \frac{225}{306} = 0.735$$

And according to geometric distribution 2nd version (with 0s):

Calculate p:

$$\frac{1-p}{p} = 2.8 \Rightarrow 2.8p = 1-p \Rightarrow 3.8p = 1 \Rightarrow p = 0.263$$

calculate $P(2\text{ positive nodes}|\text{ survived})$:

$$PMF(x=2) = (1-p)^2 p = (1-0.263)^2 \cdot 0.263 = 0.143$$

So, the probability that the patient will survive within 5 years is:

$$\boldsymbol{P = \frac{0.143 \cdot 0.735}{0.143 \cdot 0.735 + 0.092 \cdot 0.265} = \frac{0.105}{0.105 + 0.024} = \frac{0.105}{0.129} = 0.814}$$

Likelihood

1. The likelihood function:

$$L(\theta|x_n) = \prod_{i=1}^{n} \frac{2x_i}{\theta} e^{\frac{-x_i^2}{\theta}} = \frac{2^n}{\theta^n} \cdot \prod_{i=1}^{n} x_i \cdot e^{\frac{-x_i^2}{\theta}}$$

Log likelihood:

$$\ln L(\theta|x_n) = \ln 2^n + \ln \theta^{-n} + \ln \prod_{i=1}^{n} x_i + \ln \prod_{i=1}^{n} e^{\frac{-x_i^2}{\theta}}$$

$$= n \ln 2 - n \ln \theta + \sum_{i=1}^{n} \ln (x_i) + \sum_{i=1}^{n} \ln \left(e^{\frac{-x_i^2}{\theta}} \right)$$

$$= n \ln 2 - n \ln \theta + \sum_{i=1}^{n} \ln (x_i) + \sum_{i=1}^{n} \frac{-x_i^2}{\theta} \ln(e)$$

$$= n \ln 2 - n \ln \theta + \sum_{i=1}^{n} \ln (x_i) + \sum_{i=1}^{n} \frac{-x_i^2}{\theta} \cdot 1$$

$$= n \ln 2 - n \ln \theta + \sum_{i=1}^{n} \ln (x_i) - \frac{1}{\theta} \sum_{i=1}^{n} x_i^2$$

The derivative of LL with respect to θ :

$$\frac{\partial \ell}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i^2$$

To find θ that maximizes LL we will equal the derivative to zero:

$$\frac{\partial \ell}{\partial \theta} = 0 \Rightarrow -\frac{n}{\hat{\theta}} + \frac{1}{\hat{\theta}^2} \sum_{i=1}^{n} x_i^2 = 0$$

$$\frac{-n\hat{\theta} + \sum_{i=1}^{n} x_i^2}{\hat{\theta}^2} = 0$$

$\theta > 0$, *So:*

$$-n\hat{\theta} + \sum_{i=1}^{n} x_i^2 = 0$$

$$\boldsymbol{\hat{\theta} = \frac{\sum_{i=1}^{n} x_i^2}{n}}$$

2. Estimate θ from the sample $(0.5,0.5,1)$:

$$\boldsymbol{\hat{\theta} = \frac{(0.5^2 + 0.5^2 + 1^2)}{3} = 0.5}$$

Hypothesis

We need to test:

$H_0: p_x = p_y$ versus $H_0: p_x > p_y$

$$\bar{X} \sim N\left(p_x, \frac{p_x(1-p_x)}{n_x}\right) \qquad \bar{Y} \sim N\left(p_y, \frac{p_y(1-p_y)}{n_y}\right)$$

$$\bar{X} - \bar{Y} \sim N\left(p_x - p_y, \frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}\right)$$

Since sample sizes are “large enough” we can use normal approximation.

The test statistic:

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}}$$

Where:

$n_x = 100$

$n_y = 150$

$$\hat{p}_x = \frac{60}{100} = 0.6$$

$$\hat{p}_y = \frac{70}{150} = 0.467$$

So:

$$z = \frac{0.6 - 0.467}{\sqrt{\frac{0.6 \cdot (1 - 0.6)}{100} + \frac{0.467 \cdot (1 - 0.467)}{150}}} = \frac{0.133}{\sqrt{\frac{0.24}{100} + \frac{0.249}{150}}} = 2.087$$

For 1% significance level we will reject H_0 for $z > 2.326$ (Upper-Tailed Test).

Since $\mathbf{2.087 < 2.326}$ we cannot reject $\boldsymbol{H_0}$ that $\boldsymbol{p_x = p_y}$.