CS2: AMI mortality

\[ n = 403 \text{ patients} \]
\[ s = 72 \text{ died} \]
in 1st 30 days

 Neyman:
\[ \hat{\theta} = \frac{\bar{Y}}{n} \text{ is reasonable estimate of } \theta \text{ for population mean} \]

The entire diagram can be expressed in math as follows:

\[ Y_i \sim \text{Bernoulli}(\theta) \]

\[ \sum_{i=1}^{n} Y_i \text{ is } \bar{Y} \]

unknown constant

\( Y_1, Y_2, \ldots, Y_n \text{ are random variables (sampling distribution)} \)

Frequentist model in CS2
Lecture notes
Fisher (1925) (2)

part 8

Given \( \theta \), write down marginal dist. (PMF or PDF) for \( Y_i \):
\[
p(Y_i = y_i; \theta) = P(\Theta_i = \theta, Y_i = y_i)
\]

For \( y_i = 0 \) or \( 1 \):
\[
p(Y_i = 1) = \theta \cdot (1 - \theta) = (1 - \theta) \cdot \theta
\]

For \( y_i = 0 \):
\[
p(Y_i = 0) = \frac{\theta}{\theta}
\]

(1) write down joint dist. of

\( Y = (Y_1, ..., Y_n) \)

Joint = product of marginals

\[
p(Y = y; \theta) = p(Y_1 = y_1, ..., Y_n = y_n)
\]

\[
= \prod_{i=1}^{n} p(Y_i = y_i; \theta)
\]

Compute science
Comment: unrealistic to pretend θ is constant for all ω1ω.

\[ (\omega_1 : θ_i) \text{ is Bernoulli } (θ_i) \]

\[
\begin{align*}
\text{# data values} &= n \\
\text{# unknowns} &= n
\end{align*}
\]

\[ \hat{θ}_6 = 1 \]

\[ P_\text{true}(y|θ) = \prod_{i=1}^{n} P_\text{true}(y_i|θ) \]

\[ = \prod_{i=1}^{n} \left[ θ_i^{y_i} (1-θ_i)^{1-y_i} \right] = \]

\[ \left[ θ^1(1-θ)^{1-7} \right] \cdot \left[ θ^2(1-θ)^{1-7} \right] \cdot \ldots \cdot \]

\[ \left[ θ^n(1-θ)^{1-7} \right] \]
$\sum \gamma_i + \gamma_j \cdots + \gamma_k (1 - \theta) + (1 - \gamma_i) + \cdots + (1 - \gamma_k) = \sum_{i=1}^{n} \gamma_i = 2 \gamma_i$

\[ \theta^5 (1 - \theta)^{n - 5} \]

$\text{Laplace/Fisher}$

$p_{\theta}(x | \theta)$ looks like a function of $\theta$ for fixed $\theta$; but let's just reverse our thinking: regard it instead as a function of $\theta$ for fixed $x$:

\[ \mathcal{L}(\theta | x) = \int p_{\theta}(x | \theta) \, d\theta \]

(Re) Likelihood function for $\theta$ (c real $\to$)
Here
\[ \ell(\theta | x) = c \theta^5 (1-\theta)^{n-5} \]

Def.: 
\[ \ell(\theta | s) = \frac{s(x)}{27} \]

If \( \ell(\theta | x) \) depends on \( x \) only through some function \( s(x) \), \( s(x) \) is said to be sufficient for \( \theta \), \( s \) is a sufficient statistic for \( \theta \).

Suppose dimensionality-reduction in data space from \( n \) to 1

\[\dim(x) = n\]
\[\dim(\theta) = 1\]
\[\dim(s) = 1\]

\[\dim(s) = \dim(\theta)\]
uncertainty about sampling

Fisher claimed: no model nec.
in practice: model "walks in
usually door with client": FALSE

\[
\text{total info in } \mathcal{X} \ni \text{info in } \mathcal{X}_t \text{ to check whether your sampling dist is good} \quad \text{not diagnostic info}
\]
\[ x_1 = 1011000101 + 5 = 5 \\
\[ x_2 = 1111100000 + 5 = 5 \\
\[ x_3 = 1010101010 - 5 = 5 \\
\[ x_4 = 1011000101 - 5 = 5 \\

**Fisher: trust me:**

Find \( \theta \) that maximize \( \ell(\theta | x) \):

\[ \theta_{MLE} = \arg \max_{\theta \in (0,1)} \ell(\theta | x) \]

Maximum likelihood estimate

\[ \frac{d}{d\theta} \ell(\theta | x) = \frac{5}{n} \left( \theta^5 (1-\theta)^{n-5} \right) \]
Def. \( \log \text{likelihood} \) here

\[ \ell(\theta | x) = \ell(\theta | x) \]

\[ = -5 \log \theta + c_{n-5}(1 - \theta) \]

to find \( \theta \) once, can diff. either

\[ \ell(\theta | x) \lor \ell(\theta | x) \& \text{set} \]

\[ \log x \]

strictly increasing (inequality property)
\[
\frac{1}{\theta} P(\theta | 15) = \frac{5}{\theta} + \frac{5-5}{1-\theta} (-1)
\]

\[
= \frac{5}{\theta} - \frac{5}{1-\theta}
\]

\[
= \frac{5(1-\theta) - \theta(5-5)}{\theta(1-\theta)}
\]

\[
= \frac{5 - 5\theta - 5\theta + \theta^2}{\theta(1-\theta)} = 0
\]

\[
\text{If } \theta = \hat{\theta}, \text{ then } \text{MLE} = \frac{5}{n} = \frac{5}{22} (\text{Maximum})
\]

\[
= \frac{22}{462} \approx 0.179
\]
\[ SE_{\theta_{\text{MLE}}} = ? \]

here:

\[ \hat{\theta}_{\text{MLE}} = 7 \]

\[ V(\hat{\theta}_{\text{MLE}}) = \frac{\theta(1-\theta)}{n} \]

\[ \therefore SE_{\text{RS}}(\hat{\theta}_{\text{MLE}}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \]

\[ = 0.019 \]

**High uncertainty** = **Low information**

\[ \text{high variance of } \hat{\theta}_{\text{MLE}} \]

\[ \text{about } \theta \text{ from } \hat{\theta}_{\text{MLE}} \]
Fisher Information of a \( \theta \) MLE:

\[
I(\hat{\theta}_{\text{MLE}}) = \left[ -\frac{\partial^2 \log f(\theta_{\text{MLE}})}{\partial \theta^2} \right]_{\hat{\theta}}
\]

Here

\[
I(\hat{\theta}_{\text{MLE}}) = \frac{n}{\theta(1-\theta)}
\]

Fisher showed \( I(\hat{\theta}_{\text{MLE}}) = O(n) \)

\[
V_{rs}(\hat{\theta}_{\text{MLE}}) = \left[ I(\hat{\theta}_{\text{MLE}}) \right]^{-1}
\]

So

\[
\text{SE}_{rs}(\hat{\theta}_{\text{MLE}}) = \sqrt{V_{rs}(\hat{\theta}_{\text{MLE}})}
\]
\[ SE_{RS}(\hat{\theta}_{\text{MUE}}) = \sqrt{\hat{I}(\hat{\theta}_{\text{MUE}})} \]

Fisher showed 3 things:

1. \[ E_{RS}(\hat{\theta}_{\text{MUE}}) = \theta + O\left(\frac{1}{n}\right) \]
   - i.e., MUE may be biased but
     \[ \text{bias} \to 0 \text{ at } \frac{1}{n} \text{ rate (fast)} \]

2. \[ SE_{RS}(\hat{\theta}_{\text{MUE}}) \approx \sqrt{\hat{I}(\hat{\theta}_{\text{MUE}})} \]

3. For large \[ \frac{1}{n} \]
   \[ \hat{\theta}_{\text{MUE}} \sim N(\theta, \frac{1}{n}) \]
\[ \hat{\theta} \pm 1.96 \cdot \text{SE}(\hat{\theta}_{\text{MLE}}) \]

- 95% likelihood-based interval for \( \theta \):

\[ \hat{\theta}_{\text{MLE}} \pm 1.96 \cdot \text{SE}(\hat{\theta}_{\text{RS MSE}}) \]