Bayesian example

Inference

Random sample of $n = 112$ medical records of heart attack patients, 4 of them had an unplanned transfer to the Intensive Care Unit (ICU).

Goal:

Estimate $\theta$ from the dataset $Y = (y_1, \ldots, y_n)$, where $y_i = \{1 \text{ if ICU, } 0 \text{ otherwise}\}$ and construct a 95% interval for $\theta$ using the Bayesian approach to inference.

For comparison: 95% frequentist interval from two
points of view: (Neyman, Central Limit Theorem) \[ \hat{\theta} = \frac{5}{112} = 0.044 \]

SE(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{(0.05)(1-0.05)}{112}} = 0.0357

approximate (large-sample) 95\% confidence interval \[ n: \hat{\theta} \pm 1.96 \times \text{SE}(\hat{\theta}) \]

\[ 0.044 \pm 1.96 \times 0.0357 \]

\[ (0.001, 0.070) \]

\( \hat{\theta}_{\text{MLE}} = \frac{5}{112} = 0.044 \)

\[ \text{SE}(\hat{\theta}_{\text{MLE}}) = \sqrt{\text{I}^{-1}(\hat{\theta}_{\text{MLE}})} \]

likelihood function

\[ 2(\theta | y) = \theta^5(1-\theta)^{n-5} \]

log likelihood function

\[ \log 2(\theta | y) = 5 \log \theta + (n-5) \log (1-\theta) \]
\[ \frac{d}{d\theta} \ln(p(\theta)) = \frac{5}{\theta} - \frac{n-5}{1-\theta} = \frac{5 - n\theta}{\theta(1-\theta)} \]

\[ = 0 \text{ if } \theta = \frac{5}{n} \]

\[ \text{MLE} = \frac{5}{n} \int \text{Calculating the Fisher information} \]

\[ - \frac{d^2}{d\theta^2} \ln(p(\theta)) \bigg|_{\theta = \frac{5}{n}} = \frac{5}{\theta^2} (1-\theta)^2, \text{ so} \]

\[ - \frac{d^2}{d\theta^2} \ln(p(\theta)) \bigg|_{\theta = \frac{5}{n}} = \frac{1}{n} \]

\[ \text{after simplification} \]

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

\[ \text{So } SE(\hat{\theta} \text{ MLE}) = \sqrt{\frac{\hat{\theta} \text{ MLE} (1 - \hat{\theta} \text{ MLE})}{n}} \approx \frac{\hat{\theta} \text{ MLE} (1 - \hat{\theta} \text{ MLE})}{n} \text{ (Neyman-Fisher)} \]

\[ = 0.0075 \]

\[ \text{Approximate large-sample Fisher} \]

\[ 95\% \text{ confidence interval: } \hat{\theta} \pm 1.96 \text{SE}(\hat{\theta}) \]
\( \text{Bayesian approach} \)

\[
\begin{align*}
\text{So, Neyman approx.} & \approx (\text{Fisher approx.})
\end{align*}
\]

\[
(0.01, 0.70)
\]

Reminder of Bayes's Theorem:

\( U = \text{unknown} \) (T/F)

\( D = \text{data} \) (T/F)

For problem context, provides \( P(U), P(D), P(D|U) \);

What is \( P(U|D) \)?

Definition:

\[
P(U|D) = \begin{cases} 
\frac{P(U,D)}{P(D)} & \text{if } P(D) > 0 \\
\text{undefined else}
\end{cases}
\]

So therefore:

\[
P(U,D) = P(J) \cdot P(U|J)
\]
definition \[ p(D|u) = \begin{cases} \frac{p(D,u)}{p(u)} & \text{if } p(u) \neq 0 \\ p(u) & \text{else} \end{cases} \]

so therefore

\[ p(D,u) = p(u) \cdot p(D|u) \]

but \[ p(u,D) = p(D,u) \]; therefore

\[ p(D) \cdot p(u|D) = p(u) \cdot p(D|u) \]

and \[ p(u|D) = \frac{p(u) \cdot p(D|u)}{p(D)} \]

Bayes's Theorem for TV/F propositions

In the Kaiser ICU

\[ u \leftrightarrow 0 < \theta < 1 \ (\text{unknown}) \]

\[ D \leftrightarrow y = (y_1, \ldots, y_n) \]

Case study turns out as we saw in Ams' 131, that
The analogue of Bayes's Theorem when $a$ and $b$ are not fixed propositions is obtained by simply translating the ingredients in $\mathcal{Q}$ above into unconditional and conditional probability distributions.

\[
p(\theta | y) = \frac{p(y) \cdot p(\theta) \cdot p(y | \theta)}{p(\theta) \cdot p(y | \theta)}
\]

- Posterior distribution for $\theta$
- Prior distribution for $\theta$
- Sampling distribution for $y$

Once the data vector $y$ is known, the left-hand side LHS is a function of $\theta$ for fixed $y$. 
Therefore, the right-hand side (RHS) also has to be regarded as a function of θ for fixed x.

Two consequences:

(a) \( p(y) \) is a constant (i.e., \( \theta \)), playing the role of \( \frac{1}{\text{normalizing constant}} \).

(b) \( p(\theta | y) \) is proportional to \( p(y | \theta) p(\theta) \).
Bayesians think of the sampling distribution \( p(\gamma | \theta) \)
of \( \gamma \) given \( \theta \) as a function of \( \theta \) for fixed \( \gamma \), exactly as Fisher did when he defined maximum likelihood:

\[
\ell(\theta | \gamma) = c \cdot p(\gamma | \theta)
\]

Thus \( p(\theta | \gamma) \propto p(\theta) \cdot \ell(\theta | \gamma) \)

for \( \theta \) given \( \gamma \).
Now we already know in the Kaiser ICU care study that
\[ P(\theta | z) = \theta^5 (1-\theta)^n, \]
so
\[ P(\theta | z) \propto P(\theta) \cdot \theta^5 (1-\theta)^n. \]

What shall we use for the prior distribution \( p(\theta) \)?

\( p(\theta) \) quantifies all information (if any) about \( \theta \) external to the dataset \( z \). Suppose (as was the case in the Kaiser example) that little was known about \( \theta \) external
to 2 other than the obvious fact that $0 < \theta < 1$. Then we want $p(\theta)$ to be a low-informative (otherwise known as diffuse) prior distribution.

It should be essentially flat across most or all of $(0,1)$.

One possibility:

$$p(\theta) = \begin{cases} 1 & \text{for } 0 < \theta < 1 \\ 0 & \text{else} \end{cases}$$

This is the uniform distribution on $(0,1)$ if $\theta \sim \text{Uniform}(0,1)$.
And generally, it would be nice to find a family of distributions on $(0,1)$ that made calculating the posterior distribution easy.

Q: Is there such a family?

A: Yes: it's called the family of Beta distributions $\text{Beta}(\alpha, \beta)$, indexed by the two quantities $(\alpha > 0, \beta > 0)$:

\[ p(\theta) = c \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \]

Note: $\text{Beta}(1, 1) = \text{Uniform}(0, 1)$.
Recall that $p(\theta | x) \propto p(x | \theta) \cdot p(\theta)$

with $p(\theta) = c \theta^{\alpha-1} (1-\theta)^{\beta-1}$, the posterior distribution becomes $p(\theta | x) \propto \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}$

In other words, if the prior is Beta($\alpha$, $\beta$) then the posterior is Beta($\alpha+x$, $\beta+n-x$) when the prior and posterior are different members of the same family of probability distributions.
The prior is said to be conjugate to the likelihood function.

Thus we have our first conjugate updating rule:

\[ \Pr(\theta | y) \sim \text{Beta}(\alpha + s, \beta + n - s) \]

R demo: plot Uniform(0, 1) = Beta(1, 1) prior and Beta(\alpha + s, \beta + n - s) = Beta(1 + 4, 1 + 112 - 4) = Beta(5, 109) posterior on same scale.
Use $\text{beta}(0.025, 5, 109)$ and $\text{beta}(0.975, 5, 109)$ to compute 95\% Bayesian interval & compare with frequentist 95\% CI (0.01, 0.07).

$\text{beta}(0.025, 5, 109) = 0.0145$

$\text{beta}(0.975, 5, 109) = 0.0882$

Bayes: (0.0145, 0.0882) symmetric

Naive:

Fisher: (0.00, 0.07) symmetric