STAT 431 — Applied Bayesian Analysis — Course Notes

Population Proportion: Posterior Inference

Spring 2019

Inference can include:

- estimation
- hypothesis testing
- prediction

The **frequentist** approach bases all inference on the data and its distribution, regarding the parameter(s) as *fixed*.

The Bayesian approach treats the parameter(s) as random.

Let's compare them in the case of a binomial proportion ...

Frequentist Methods

- Point estimate:
$$\hat{\pi} \;=\; \frac{y}{n}$$
(which is a MOM estimate and a MLE — see Cowles)

  Eg: survey — $\quad y = 12$ out of $n = 70$ own pets
$$\hat{\pi} \;=\; \frac{12}{70} \;\approx\; 0.171$$

- Standard error (estimated):
$$\mathrm{se}(\hat{\pi}) \;=\; \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \;\approx\; 0.045$$

- Confidence Interval:

  The "Wald" $(1-\alpha)100\%$ CI is

$$\hat{\pi} \quad \pm \quad z_{\alpha/2}\,\mathrm{se}(\hat{\pi})$$

  For our survey data, the 95% interval is

$$\approx \quad (0.08,\, 0.26)$$

  Q: Would a frequentist statistician say that π has a 95% approx. probability of being in this computed interval?

- Hypothesis Test:

$$\text{e.g.} \qquad H_0 : \pi \geq \pi_* \qquad H_1 : \pi < \pi_*$$

Wald statistic:

$$z = \frac{\hat{\pi} - \pi_*}{\mathrm{se}(\hat{\pi})}$$

For H_0 : at least 30% have pets, we get

$$z \approx \frac{0.171 - 0.3}{0.045} \approx -2.85$$

which is significant for rejecting H_0.

R Example 4.1:

Population Proportion — Frequentist Methods

Bayesian Methods

Basic idea: Use the posterior distribution for everything.

Eg: Recall $\mathrm{beta}(\alpha, \beta)$ prior example, which led to

$$\pi \mid Y = 12 \quad \sim \quad \mathrm{beta}(12 + \alpha,\, 58 + \beta)$$

- Point Estimate:

  Usually the **posterior mean**:

$$\mathrm{E}(\pi \mid Y = y) \;=\; \mathrm{E}(\pi \mid y)$$

  (but could alternatively use posterior median or mode)

  Eg: For beta prior (see Cowles, Table A.2)

$$\mathrm{E}(\pi \mid Y = 12) \;=\; \frac{12+\alpha}{(12+\alpha)+(58+\beta)} \;=\; \frac{12+\alpha}{70+\alpha+\beta}$$

  so for the uniform $(\alpha = \beta = 1)$

$$\mathrm{E}(\pi \mid Y = 12) \;=\; \frac{12+1}{70+2} \;\approx\; 0.181$$

In general (binomial likelihood, beta prior):

$$\begin{aligned}
\mathrm{E}(\pi \mid y) &= \frac{\alpha + y}{\alpha + \beta + n} \\
&= \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \cdot \frac{y}{n} \\
&= \lambda \cdot \underbrace{\mathrm{E}(\pi)}_{\text{prior mean}} + (1 - \lambda) \underbrace{\hat{\pi}}_{\substack{\text{sample} \\ \text{proportion}}}
\end{aligned}$$

Note: Posterior mean is sample proportion "shrunk" toward the prior mean.

Q: As $n \to \infty$, what happens to λ? What happens to the posterior mean? Does the prior becomes less important or more important?

Also, can show that

$$\text{posterior mode} = \frac{y+\alpha-1}{(y+\alpha)+(n-y+\beta)-2}$$

which for the uniform prior ($\alpha = \beta = 1$) gives

$$\frac{y}{n}$$

(In general, the posterior mode under a *flat* prior should be the same as the maximum likelihood estimate.)

- Posterior Standard Deviation:

  Bayesian analogue of the standard error:

$$\sqrt{\mathrm{Var}(\pi \mid y)}$$

  Eg: For beta prior (see Cowles, Table A.2) this is

$$\sqrt{\frac{(12+\alpha)(58+\beta)}{(12+\alpha+58+\beta)^2(12+\alpha+58+\beta+1)}}$$

  which is, for the uniform ($\alpha = \beta = 1$), approx. 0.045.

  (comparable to standard error)

Remark: For any beta prior, you can show the *frequentist* result that

$$\frac{\sqrt{\mathrm{Var}(\pi \mid Y)}}{\mathrm{se}(\hat{\pi})} \xrightarrow[n\to\infty]{} 1 \qquad (w.p.\ 1)$$

(HW?)

- Credible Interval

  A $(1-\alpha)100\%$ **(Bayesian posterior) credible interval** for a parameter θ is a statistical interval I such that

  $$\mathrm{P}(\theta \in I \mid \text{data}) \;=\; 1-\alpha$$

  [Interpret ...]

  Two main approaches:

  - equal-tailed
  - highest posterior density (HPD)

[Illustrate equal-tailed and HPD ...]

R Example 4.2(a):

Population Proportion — Credible Intervals

- Posterior Probabilities and Testing

Given

$$H_0 : \theta \in \Theta_0 \qquad\qquad H_1 : \theta \in \Theta_1$$

a Bayesian can assign posterior probabilities

$$\begin{aligned} \mathrm{P}(H_0 \mid \textsf{data}) &= \mathrm{P}(\theta \in \Theta_0 \mid \textsf{data}) \\ \mathrm{P}(H_1 \mid \textsf{data}) &= \mathrm{P}(\theta \in \Theta_1 \mid \textsf{data}) \end{aligned}$$

Eg: A Bayesian can assign probabilities to

$$H_0 : \pi \geq 0.3 \qquad\qquad H_1 : \pi < 0.3$$

R Example 4.2(b):

Population Proportion — Posterior Probabilities

Posterior Predictive Distributions

Let

$$\begin{aligned} \theta &= \text{the model parameter} \\ y &= \text{the observed data} \\ Y^* &= \text{unobserved (new) data} \end{aligned}$$

Suppose θ has continuous posterior density $p(\theta \mid y)$.

The **posterior predictive distribution** for Y^* is defined by the density

$$p(y^* \mid y) = \int \underbrace{p(y^* \mid \theta)}_{\substack{\text{model for} \\ \text{new data}}} \underbrace{p(\theta \mid y)}_{\substack{\text{posterior} \\ \text{from obs. data}}} d\theta$$

To derive this, we assume Y^* is conditionally independent of data Y, given θ: $\quad p(y^* \mid \theta, y) \;=\; p(y^* \mid \theta)$

Then

$$
\begin{aligned}
p(y^* \mid y) &= \int p(y^*, \theta \mid y)\, d\theta \\
&= \int p(y^* \mid \theta, y)\, p(\theta \mid y)\, d\theta \\
&= \int p(y^* \mid \theta)\, p(\theta \mid y)\, d\theta
\end{aligned}
$$

Eg: Pet survey (binomial data, $\theta = \pi$)

Suppose we plan to survey $n^* = 10$ new people. Let

$$Y^* \quad = \quad \text{number of them having pets}$$

Then

$$Y^* \mid \pi \quad \sim \quad \text{binomial}(10, \pi)$$

Let's use our posterior from the uniform prior ($\alpha = \beta = 1$):

$$\pi \mid Y = 12 \quad \sim \quad \text{beta}(13, 59)$$

The posterior predictive distribution for Y^* has density (for $y^* = 0, 1, \ldots 10$)

$$
\begin{aligned}
p(y^* \mid Y = 12) &= \int_0^1 \binom{10}{y^*} \pi^{y^*} (1-\pi)^{10-y^*} \\
&\qquad \cdot \frac{\Gamma(72)}{\Gamma(13)\Gamma(59)} \pi^{12} (1-\pi)^{58} \, d\pi \\
&= \binom{10}{y^*} \frac{\Gamma(72)}{\Gamma(13)\Gamma(59)} \\
&\qquad \cdot \int_0^1 \underbrace{\pi^{y^*+12} (1-\pi)^{68-y^*}}_{\text{kernel of } \mathrm{beta}(y^*+13,\, 69-y^*)} \, d\pi \\
&= \cdots
\end{aligned}
$$

$$
\begin{aligned}
\cdots \quad &= \quad \binom{10}{y^*} \frac{\Gamma(72)}{\Gamma(13)\Gamma(59)} \frac{\Gamma(y^*+13)\Gamma(69-y^*)}{\Gamma(82)} \\
&= \quad \frac{\binom{10}{y^*} \, 71 \, \binom{70}{12}}{81 \binom{80}{y^*+12}} \qquad \text{for } y^* = 0, 1, \ldots 10
\end{aligned}
$$

(a type of *beta-binomial* distribution)

R Example 4.3:

Population Proportion — Posterior Predictive Distn.