STAT 431 — Applied Bayesian Analysis — Course Notes

Population Proportion: Posterior Inference

Fall 2022

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:

$$\hat{\theta} = \frac{y}{n}$$

(which happens to be a MOM estimate and a MLE)

Eg: survey — $y = 12$ out of $n = 70$ own pets

$$\hat{\theta} = \frac{12}{70} \approx 0.171$$

- Standard error (estimated):

$$\mathrm{SE}(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \approx 0.045$$

- Confidence Interval:

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

  $$\hat{\theta} \;\pm\; z_{\alpha/2}\,\mathrm{SE}(\hat{\theta})$$

  For our survey data, the 95% interval is

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

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

- Hypothesis Test:

$$\text{e.g.} \qquad H_0:\ \theta \geq \theta_* \qquad H_1:\ \theta < \theta_*$$

Wald statistic:

$$z \;=\; \frac{\hat{\theta} - \theta_*}{\mathrm{SE}(\hat{\theta})}$$

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 1.2:

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

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

- Point Estimate:

  Usually the **posterior mean**:

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

  (but could alternatively use posterior median or mode)

  Eg: For beta prior (see BSM Appendix A.1)

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

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

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

In general (binomial likelihood, beta prior):

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

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

Q: As $n \to \infty$, what happens to w_n? 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.)

A posterior mode is also called a **maximum a posteriori (MAP)** estimator.

- Posterior Standard Deviation:

  Bayesian analogue of the standard error:

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

  Eg: For beta prior (see BSM Appendix A.1) 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

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

(HW?)

- Credible Interval

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

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

  [Interpret ...]

  Two main approaches:

  - equal-tailed
  - highest posterior density (HPD)

[Illustrate equal-tailed and HPD ...]

R Example 1.3(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

$$\begin{aligned} \text{Prob}(H_0 \mid \mathsf{data}) &= \text{Prob}(\theta \in \Theta_0 \mid \mathsf{data}) \\ \text{Prob}(H_1 \mid \mathsf{data}) &= \text{Prob}(\theta \in \Theta_1 \mid \mathsf{data}) \end{aligned}$$

Eg: A Bayesian can assign probabilities to

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

R Example 1.3(b):

Population Proportion — Posterior Probabilities

Posterior Predictive Distributions

Let

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

Suppose $\boldsymbol{\theta}$ has continuous posterior density $p(\boldsymbol{\theta} \mid \boldsymbol{y})$.

The **posterior predictive distribution (PPD)** for $\boldsymbol{Y}^*$ is defined by the density

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

To derive this, we assume $\boldsymbol{Y}^*$ is conditionally independent of data $\boldsymbol{Y}$, given $\boldsymbol{\theta}$:

$$f(\boldsymbol{y}^* \mid \boldsymbol{\theta}, \boldsymbol{y}) \;=\; f(\boldsymbol{y}^* \mid \boldsymbol{\theta})$$

Then

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

Eg: Pet survey (binomial data)

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

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

Then

$$Y^* \mid \theta \quad \sim \quad \mathrm{Binomial}(10, \theta)$$

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

$$\theta \mid Y = 12 \quad \sim \quad \mathrm{Beta}(13, 59)$$

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

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

$$\cdots = \binom{10}{y^*} \frac{\Gamma(72)}{\Gamma(13)\Gamma(59)} \frac{\Gamma(y^*+13)\Gamma(69-y^*)}{\Gamma(82)}$$

$$= \frac{\binom{10}{y^*} 71 \binom{70}{12}}{81 \binom{80}{y^*+12}} \qquad \text{for } y^* = 0, 1, \ldots 10$$

(corresponding to a type of *beta-binomial* distribution)

R Example 1.4:

Population Proportion — Posterior Predictive Distn.