Bayesian Statistics

One-parameter models

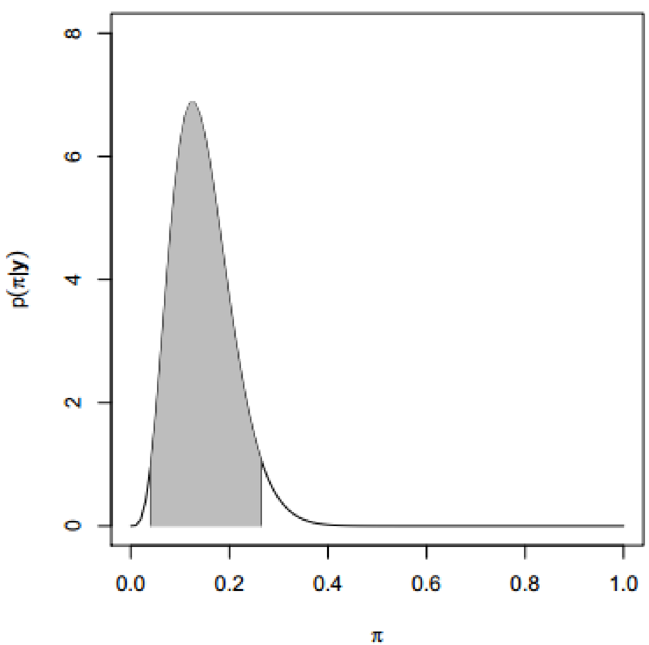
Nan Lin
Department of Mathematics
Washington University in St. Louis

$y|\theta \sim Binomial(n, \theta)$

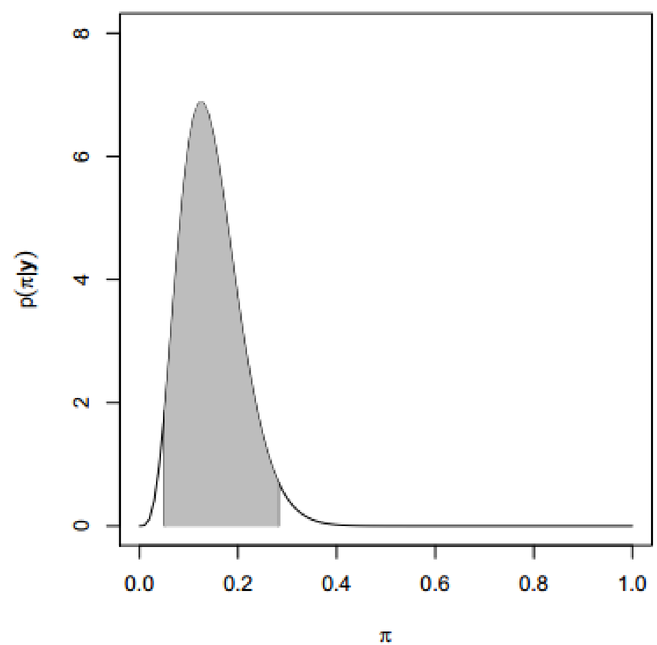
- Prior: $\theta \sim U(0,1)$
- Posterior distribution: $\theta|y \sim beta(y+1, n-y+1)$
- Predictive distribution: $\tilde{y}|y \sim Bernoulli(\frac{y+1}{n+2})$
- Summary of posterior distribution
  - Posterior mean: $\mathrm{E}(\theta|y) = \frac{y+1}{n+2}$
  - Posterior median, mode
  - Posterior variance: $var(\theta|y) = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$
  - Posterior quantile: find θ_τ such that $P(\theta \le \theta_\tau|y) = \tau$

Summary of posterior distribution (cont)

- $100(1-\alpha)\%$ credible (posterior) interval $[a, b]$:
$$P(\theta \in [a,b] | y) = 1 - \alpha$$
  - Highest posterior density (HPD) region
  - Symmetric (quantile) posterior interval: $[\theta_{\frac{\alpha}{2}}, \theta_{1-\frac{\alpha}{2}}]$



(a) 95% HPD



(b) Symmetric 95%

Credible interval

- If the posterior distribution is not unimodal,
  - HPD interval comprises disjoint intervals
  - Symmetric (quantile) posterior interval is still a continuous interval
  - However, in such a situation, using a single interval is probably not a good idea

What if a different prior is used?

- Let's consider the following prior

$$\theta \sim beta(\alpha, \beta)$$

  - $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$
  - Hyperparameters: α, β
- Posterior density:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

- Posterior distribution:

$$\theta|y \sim beta(y+\alpha, \beta+n-y)$$

- Posterior mean:

$$E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n}$$

- Posterior variance:

$$var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} = \frac{E(\theta|y)[1-E(\theta|y)]}{\alpha+\beta+n+1}$$

Comments

- Posterior mean is a weighted average of the posterior mean and sample mean
  - Prior mean: $E(\theta) = \frac{\alpha}{\alpha+\beta}$
  - Sample mean (MLE): y/n
  - $E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n} = wE(\theta) + (1-w)y/n$, where $w = \frac{\alpha+\beta}{\alpha+\beta+n}$
- As $n \to \infty, w \to 0$, and $E(\theta|y) \to \mathrm{y/n}$
  - The influence of prior distribution becomes negligible for large sample
- As $n \to \infty$, one can show that $\theta|y$ can be well approximated by a normal distribution
  - Different parameterization lead to different approximation accuracy
  - For example, $logit(\theta) = \log[\theta/(1-\theta)]$ will make the normal approximation more accurate

$y|\theta \sim N(\theta, \sigma^2)$

- Prior: $\theta \sim N(\mu_0, \tau_0^2)$
- Likelihood: $p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}$
- Posterior density: $p(\theta|y) \propto \exp\left(-\frac{(\theta-\mu_1)^2}{2\tau_1^2}\right)$, where

  $\mu_1 = \frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}$ and $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$

  - Precision: $1/variance$
- Posterior distribution: $\theta|y \sim N(\mu_1, \tau_1^2)$
- Predictive distribution: $\tilde{y}|y \sim N(\mu_1, \sigma^2 + \tau_1^2)$

$y_1, \ldots, y_n|\theta \sim N(\theta, \sigma^2)$

- Prior: $\theta \sim N(\mu_0, \tau_0^2)$
- Likelihood: $p(y_1, \ldots, y_n|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\theta)^2}{2\sigma^2}}$
- Posterior distribution:
- Posterior mean: