Bayesian Inference

STAT 211 - 509

2018-10-22

Bayes Theorem

Let X and Y be two discrete random variables:

- marginal pmfs f_X and f_Y (probabilities of individual values):
  - $P(X = x) = f_X(x)$
  - $P(Y = y) = f_Y(y)$
- conditional pmf of Y, given the value of X
- $P(Y = y|X = x) = f_{Y|X}(y|x)$

  Bayes Theorem:

$$P(X = x|Y = y) = \frac{P(Y = y|X = x)P(X = x)}{\sum_x P(Y = y|X = x)P(X = x)}$$

Posterior Distributions

Suppose we want to estimate the unknown parameter of some distribution f. We obtain a random (IID) sample of observations (*data*) from f, resulting in a likelihood of:

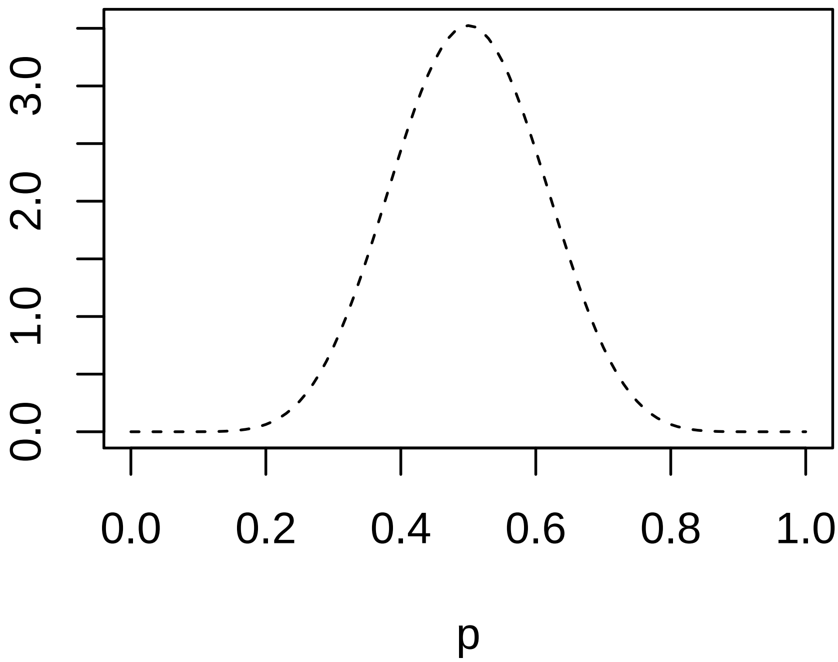
$$L(\theta) = f(data|\theta)$$

Suppose we are willing to treat the parameter θ, a fixed quantity in the population, nevertheless as if it were a random variable, with **prior** distribution g. Then an alternative strategy for inference could be based on an application of Bayes Theorem:

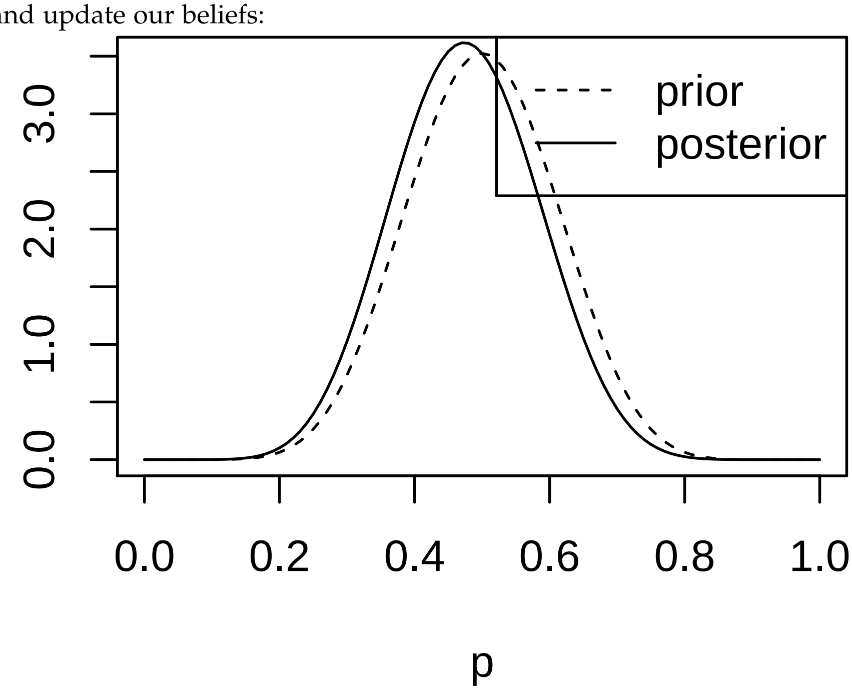
$$\begin{aligned}
\text{posterior} &= h(\theta|data) \\
&= \frac{f(data|\theta)g(\theta)}{\sum_\theta f(data|\theta)g(\theta)} \\
&= \frac{\text{likelihood} \times \text{prior}}{\sum \text{likelihood} \times \text{prior}}
\end{aligned}$$

Illustration

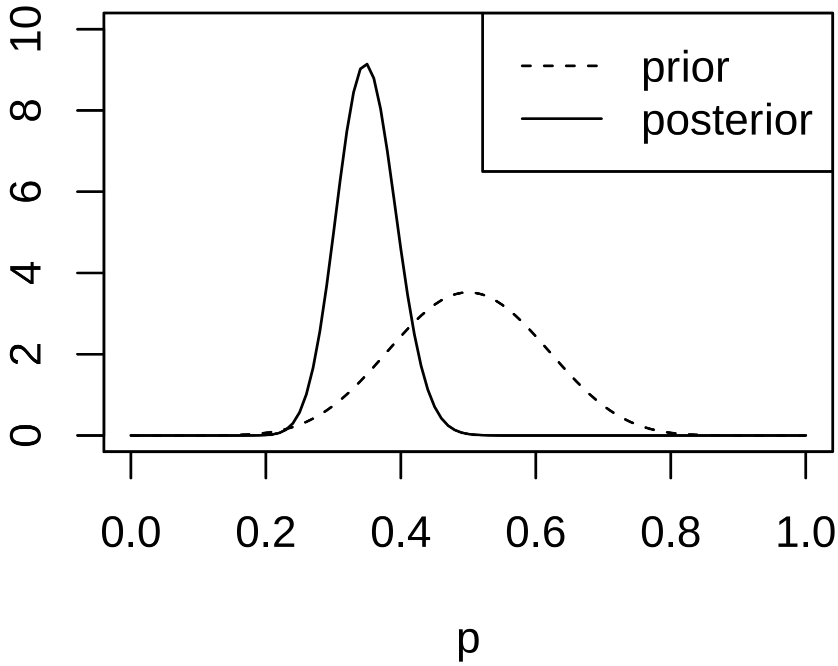
We have prior beliefs about the vote share p a candidate will recieve in an election:



We poll one voter in their district to get a sense of voter intention, and update our beliefs:



We poll many voters and use the combined information to update our beliefs:



Example: Billiards

- See billiards.R:
- Prior to playing any games, how confident are you that you are the better player in this matchup?
- After winning the first 3 games in a row, now how confident are you?
- After then losing the next 2 games (for a total record of 3 wins and 2 losses), now how confident are you?

.pull-left[Suppose you meet a colleague to discuss business over games of billiards. Suppose you have never played billiards with this particular person before and do not have much confidence in your billiards skills. Let p be the probability that you win any one game with your colleague. Going into your first game, suppose your "prior" probabilities for the value of p, call them $g(p)$, are as given in Table 2.]

.pull-right[

p	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$g(p)$	0.00	0.10	0.25	0.20	0.15	0.10	0.05	0.05	0.05	0.05	0.00

]

Frequentist vs. Bayesian Inference

- Frequentist:
  - Probability refers to limiting relative frequencies.
  - Parameters are fixed, unknown constants.

  - Statistical procedures designed to have long-run frequency properties.

- Bayesian:

  - Probability described degree of belief, not limiting frequency.
  - We can make probability statements about parameters.
  - We make inferences about a parameter by producing a probability distribution for it.

Conjugate Priors

- When prior and posterior are in same family, the prior is said to be *conjugate* with respect to the model.
- With conjugate priors, can easily draw samples directly from posterior.
- Example:

  - Binomial likelihood and beta prior: posterior also binomial
  - Normal likelihood and normal prior: posterior also normal

Functions of Parameters

- How to make inference about a function $\tau = g(\theta)$
- Recall how we solved the problem when the density of X was given as f_X and we found out density for $Y = g$. We will apply the same reasoning here.
- The posterior CDF for τ is

$$H(\tau|x^n) = P(g(\theta) \leq \tau) = \int_A f(\theta|x^n)d\theta$$

Where $A = \{\theta : g(\theta) \leq \tau\}$

- The posterior density is

$$h(\tau|x^n) = H'(\tau|x^n)$$

Simulation

- The posterior can be approximated by simulation
- If we draw $\theta_1, \ldots, \theta_B \sim p(\theta|x^n)$, then a histogram of $\theta_1, \ldots, \theta_B$ approximates the density $p(\theta|x^n)$
- Mean $\bar{\theta}_n = E(\theta|x^n)$ is
$$\frac{1}{B}\sum_{j=1}^{B}\theta_j$$
- Let $\tau_i = g(\theta_i)$, then $\tau_1, \ldots, \tau_B$ is a sample from $f(\tau|x^n)$
- This avoids the need for any analytical calculation.

Credible intervals

- Once we have $p(\theta|X)$ we can create intervals into which θ falls with a certain probability–a **credible interval**.
- Unlike a confidence interval, for a $1-\alpha$ credible interval we can say that with probability $1-\alpha$ θ falls in the interval.
- The posterior $1-\alpha$ interval can be approximated by $(\theta_{\alpha/2}, \theta_{1-\alpha/2})$ where $\theta_{\alpha/2}$ is the $\alpha/2$ sample quantile of $\theta_1, \ldots, \theta_B$

Flat Priors

- In case of a more complicated problem where there are many parameters, finding prior $f(\theta)$ seems impractical.
- An alternative is to define some sort of "noninformative prior".
- Flat prior $f(\theta) \propto \textit{constant}$ can be used as a noninformative prior.
- Flat priors are not invariant.
- Unfettered use of flat priors raises some questions

Improper Priors

- If Flat prior $f(\theta) \propto c$ where $c > 0$ is a constant, then $\int f(\theta)d\theta = \infty$
- In usual sense this is not a real probability density. Such priors are called **Improper Prior**
- The Jeffrey's rule for creating a (invariant) prior: $f(\theta) \propto I(\theta)^{1/2}$, where $I(\theta)$ is the Fisher information
- Improper priors are not a problem as long as the resulting posterior is a well defined probability distribution