27: INTRO TO BAYESIAN INFERENCE

Stat250 S25

Prof Amanda Luby

1 A Bayesian Personality Quiz

Record your answer to each question here:

Question	Q1	Q2	Q3	Q4
Answer				
Points				

2 First Bayesian Example

Example: A Des Moines register poll a few days before the 2024 Presidential election showed Kamala Harris with 51.61% of the 2-party vote share in a poll of $n = 808$ likely voters. This poll result received a lot of buzz, because in the two months leading up to this poll, this proportion was estimated between 44.7 and 47.7.

- Nov 2: 51.5% (x = 417; n=808)
- Nov 2: 44.7% (x = 358; n=800)
- Oct 2: 46.8% (x=281; n=600)
- Sept 15: 47.7% (x=382; n=800)

2.1 Overview of Bayesian Method

1. Choose (or elicit) a probability distribution to express the pre-data belief about the parameter of interest, θ.
2. Choose a model for the data given θ.
3. Observe data, $Y_1, \dots, Y_n$.
4. Update the belief about θ by combining the prior belief and the data.
5. Draw inferences using this updated belief about θ.

- **Likelihood**: a model for our data $X_1, \dots, X_n$
- X_i is drawn from a population/distribution with pdf/pmf

$$X_i \sim f(x_i|\theta)$$

- The **joint** probability model for all n data values (likelihood function)

$$f(x_1, \dots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- **Prior**: we give θ an initial probability model that reflects our beliefs about θ *prior* to observing our data
$$\theta \sim f(\theta)$$
- **Posterior**: update our prior f to reflect the information about θ that is provided by our data $x_1, \dots, x_n$
$$\theta \mid x_1, \dots, x_n \sim f(\theta \mid x_1, \dots, x_n)$$
- We compute our conditional posterior distribution using Bayes theorem:
$$f(\theta \mid x_1, \dots, x_n) = \frac{f(\theta, x_1, \dots, x_n)}{f(x_1, \dots, x_n)} = \frac{f(\theta) f(x_1, \dots, x_n \mid \theta)}{f(x_1, \dots, x_n)}$$

3 Beta-Binomial Model

- **Data** 808 respondents who responded to the 2-party presidential vote question.
-
$$f(x \mid p) = \binom{808}{x} p^x (1-p)^{808-x}$$
- **Prior**: A natural prior for a proportion is a uniform:
$$p \sim Unif[0,1] \text{ so that } f(p) = 1 \text{ for } 0 \le p \le 1$$
- **Posterior** We need to derive the following pdf for p:
$$f(p \mid x) = \frac{f(\theta) f(x \mid p)}{f(x)} = \frac{1 \times \binom{n}{x} p^x (1-p)^{n-x}}{\int_0^1 1 \times \binom{n}{x} p^x (1-p)^{n-x} dp} \text{ for } 0 \le p \le 1$$

- Which is the pdf for Beta distribution:

$$p \mid x \sim Beta(x+1, n-x+1)$$

- Iowa data: our posterior distribution for $n = 808$ and $x = 417$ is

$$p \mid x = 417 \sim Beta(418, 392)$$

```
qbeta(c(.025,.975),418,392)  # 95% credible interval for p
```

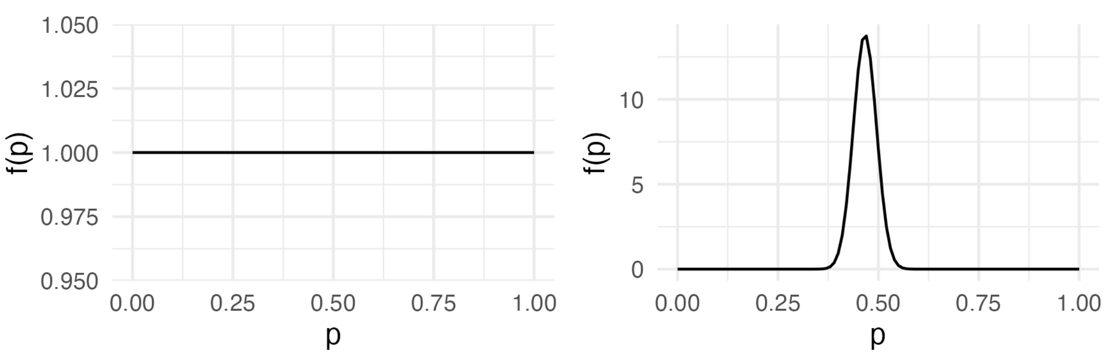
```
[1] 0.4816267 0.5503970
```

```
prop.test(417,808)$conf  # 95% confidence interval for p
```

```
[1] 0.4810195 0.5510037
attr(,"conf.level")
[1] 0.95
```

3.1 What if we chose a different prior?

In the three polls leading up to the Des Moines register poll, Harris' 2-party vote share hovered between 44.7 and 47.7 percent. What if my prior distribution took that information into account?



- A more flexible prior for p is a Beta(α, β) distribution:

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \text{ for } 0 \leq p \leq 1$$

- The posterior then looks like:

$$f(p \mid x) = \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \times \binom{n}{x} p^x (1-p)^{n-x}}{f(x)} \qquad = \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{x}}{f(x)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \text{ for } 0 \leq p \leq 1$$

- The **kernel** of the posterior (the part that involves p) looks like:

$$f(p \mid x) \propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \text{ for } 0 \leq p \leq 1$$

- Since $f(p \mid x)$ must integrate to 1 over [0,1], the kernel uniquely identifies the pdf as a Beta($x + \alpha, n - x + \beta$)
  - which means the **normalizing constant** for this kernel is

$$\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)} = \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{x}}{f(x)}$$

  - we just need to look at

$$f(\theta \mid x_1, \dots, x_n) \propto f(\theta) f(x_1, \dots, x_n \mid \theta)$$

Beta-Binomial Model

- Data: $X \mid p \sim Binom(n, p)$
- Prior: $p \sim Beta(\alpha, \beta)$
- Prior Expectation: $E(p) = \dfrac{\alpha}{\alpha+\beta}$
- Posterior: $p \mid x \sim Beta(x+\alpha, n-x+\beta)$
- Posterior Expectation: $E(p \mid x) = \dfrac{x+\alpha}{n+\alpha+\beta} = \dfrac{n}{n+\alpha+\beta}\hat{p}_{MLE} + \dfrac{\alpha+\beta}{n+\alpha+\beta}E(p)$

3.2 Comparing intervals

If we compare credible intervals for the 3 posteriors:

```
qbeta(c(.025,.975),418,392)
#> [1] 0.4816267 0.5503970
qbeta(c(.025,.975),557,551)
#> [1] 0.4732816 0.5321244
qbeta(c(.025,.975),1438,1570)
#> [1] 0.4602256 0.4959190
```

Note: on election day, Harris received 43.27% of the two-party vote share