Data 102 Lecture 6: Bayesian Inference

Topics for today:
- Review of Bayesian probability
- Forms of inference
- Start on graphical models

Bayesian modeling:

2 → Air Purifier #1
5 reviews: all positive (5 star)

40 → Air Purifier #2
20 reviews: 19 positive (5 star), 1 negative (2 star)

$\Theta_1, \Theta_2$

$\Theta_i$: probability that product $i$ gets positive review

observed actual reviews
Focus on product 1

Observe $x_1, \ldots, x_5 \in \{0, 1\}$

Unknown $\theta \in [0, 1]$; prob. pos

Likelihood

$p(x_1 | \theta) = \theta$
$p(x_1 = 0 | \theta) = 1 - \theta$

$p(x_1 = 1 | \theta) = \theta^{x_1} (1 - \theta)^{1-x_1}$

$p(x_1 = 0 | \theta) = \theta^0 (1 - \theta)^1 = \theta$

$x_1 = 1$

$x_1 = 0$

$p(x_1, \ldots, x_5 | \theta) = \theta^{x_1} (1 - \theta)^{x_2} \theta (1 - \theta)^{x_3} \ldots \theta (1 - \theta)^{x_5}$

$p(x_1, \ldots, x_5 | \theta) = \theta^{x_1 + x_2 + \ldots + x_5} (1 - \theta)^{5 - (x_1 + x_2 + \ldots + x_5)}$

$p_{\text{positive}} = \theta^{x_1 + x_2 + \ldots + x_5} (1 - \theta)^{5 - (x_1 + x_2 + \ldots + x_5)}$

$p_{\text{negative}} = \theta^{x_1 + x_2 + \ldots + x_5} (1 - \theta)^{5 - (x_1 + x_2 + \ldots + x_5)}$
\( x_1 = \begin{cases} 1: \text{Customer #1 has positive review} \\ 0: \text{Customer #1 has negative review} \end{cases} \)

\[
p(x_1 | \theta) = \begin{cases} \theta: x_1 = 1 \\ 1 - \theta: x_1 = 0 \end{cases}
\]

Bernoulli random variable = \( \theta^x (1-\theta)^{1-x} \)

---

Goal: Say something about \( \theta \)

Starting point: MLE (maximum likelihood estimation)

\[
\hat{\theta}_{MLE} = \arg\max_\theta p(x_1, \ldots, x_5 | \theta)
\]

\( x_1, \ldots, x_5 \) all equal 1

\[
= \arg\max_\theta \theta^5 (1-\theta)^0
\]

\[
= 1
\]
Product #21

\[ \hat{\theta}_{\text{MLE}} = \operatorname{argmax} \theta \quad \theta^{19} (1-\theta) \]

\[ = \operatorname{argmax} \theta \quad \log(\theta^{19} (1-\theta)) \]

\[ = \operatorname{argmax} \theta \quad 19 \log(\theta) + \log(1-\theta) \]

\[ = \frac{19}{20} \]

\[ \text{MLE: maximizer (over } \theta) \text{ of } p(x|\theta) \]

**Bayesian approach:**

\[ p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \]

\[ p(\theta|x_1,\ldots,x_5) = \frac{p(x_1,\ldots,x_5|\theta)p(\theta)}{p(x_1,\ldots,x_5)} \]

\[ \text{posterior} \quad \text{known} \quad \text{prior} \]

\[ p(x_1,\ldots,x_5) \]

\[ p(x_1,\ldots,x_5) \]
\[ p(\Theta | x_1, \ldots, x_5) \propto p(x_1, \ldots, x_5 | \Theta) \cdot p(\Theta) \]

Two questions:

1. What is prior \( p(\Theta) \)?
2. How should I make my decision?

Example prior: Beta distribution

\[ p(\Theta) \propto \Theta^{r-1} (1-\Theta)^{s-1} \quad (r, s > 0) \]

\( \text{Beta}(r, s) \)

\( \text{Beta}(1, 1) \) uniform over \([0, 1]\)

\[ \Theta^{1-1} (1-\Theta)^{1-1} = 1 \]

\( \text{Beta}(2, 1) \) skewed toward 1
\[ \text{Beta}(1, 2) \quad \text{or} \quad 1 - \theta \quad (\text{strengthen 0}) \]

"Conjugate prior" of Bernoulli

\[ p(\theta) = \theta^{r-1} (1-\theta)^{s-1} \]

\[ p(\theta | x_1, \ldots, x_5) \propto \ell(x_1, \ldots, x_5 | \theta) \quad p(\theta) \]

\[
\begin{align*}
\text{all } 1 & \quad \theta^5 \quad \theta^{r-1} (1-\theta)^{s-1} \\
& = \theta^{r+5-1} (1-\theta)^{s-1} \\
& = \text{Beta}(r+5, s)
\end{align*}
\]

Continuous data:

\[ x_1, \ldots, x_{100} \in \mathbb{R}_{\geq 0} \]

heights
$\Theta$: average height

$p(x, \theta) \propto \exp\left(\frac{1}{2\sigma^2} (x_i - \theta)^2\right)$

Gaussian with mean $\Theta$ (assume $\sigma$ is known)

$p(\theta) \propto \exp\left(\frac{1}{2\sigma^2} (\theta - \mu)^2\right)$

**Decisions:**
- Need to actually choose a $p(\theta)$ to buy.

- Ways to do this:
  - Define a loss function
  - Provide some way of
Approximate inference

Computation

- posterior mode (MAP)
- posterior mean
- $E[\theta | x]$

$\text{based on that}$

Summary of $p(\theta | x)$, decide
Inference via sampling

Lecture 8