Math 170S

6.8

Bayesian Estimation

Coin Toss

Example 6.1

Suppose there are three coins in a pocket with success (head) probabilities of 0.2, 0.5, and 0.8 respectively. I pick one coin uniformly at random and flip it twice. Calculate the probability of getting heads on both tosses.

$\mathbb{P}(\theta = 0.2) = \frac{1}{3}$ $\qquad$ $\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.2) = 0.2^2$

$\mathbb{P}(\theta = 0.5) = \frac{1}{3}$ $\qquad$ $\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.5) = (0.5)^2$

$\mathbb{P}(\theta = 0.8) = \frac{1}{3}$ $\qquad$ $\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.8) = (0.8)^2$

$$K = \mathbb{P}(X_1, X_2 = 1) = \mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.2) \cdot \mathbb{P}(\theta = 0.2) + \mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.5) \cdot \mathbb{P}(\theta = 0.5)$$

$$+ \mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.8)\, \mathbb{P}(\theta = 0.8)$$

$$= (0.2^2) \cdot \frac{1}{3} + (0.5)^2 \cdot \frac{1}{3} + (0.8)^2 \cdot \frac{1}{3} = 0.31.$$

$\pi_{prior}(\theta = 0.2) = \pi_{prior}(\theta = 0.5) = \pi_{prior}(\theta = 0.8) = \frac{1}{3}$

Coin Toss

Example 6.2

Compute the probability

$$P[\Theta = 0.2 \mid X_1, X_2 = 1]; \qquad P[\Theta = 0.5 \mid X_1, X_2 = 1];$$
$$P[\Theta = 0.8 \mid X_1, X_2 = 1].$$

data / obs.

$\pi_{post}(\theta = 0.2)$

$$= \mathbb{P}(\theta = 0.2 \mid X_1, X_2 = 1) = \frac{\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.2) \cdot \mathbb{P}(\theta = 0.2)}{\mathbb{P}(X_1, X_2 = 1)} = \frac{(0.2)^2 \cdot \frac{1}{3}}{0.31} = 0.044$$

$\pi_{post}(\theta = 0.5)$

$$= \mathbb{P}(\theta = 0.5 \mid X_1, X_2 = 1) = \frac{\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.5) \cdot \mathbb{P}(\theta = 0.5)}{\mathbb{P}(X_1, X_2 = 1)} = \frac{(0.5)^2 \cdot \frac{1}{3}}{0.31} = 0.268$$

$\pi_{post}(\theta = 0.8)$

$$= \mathbb{P}(\theta = 0.8 \mid X_1, X_2 = 1) = \frac{\mathbb{P}(X_1, X_2 = 1 \mid \theta = 0.8) \cdot \mathbb{P}(\theta = 0.8)}{\mathbb{P}(X_1, X_2 = 1)} = \frac{0.8^2 \cdot \frac{1}{3}}{0.31} = 0.688.$$

posterior pmf.

Bayesian Inference; Discrete Case

1. Let X be a random variable with distribution $f(\cdot|\theta) =: f_\theta(\cdot)$ for some parameter Θ.
2. Θ is a **discrete** random variable on Ω with an unknown pmf π.
3. **Problem:** Estimate the unknown pmf π.
4. **Input:**
   - Sample values $x_1, \ldots, x_n$ from n experiments.
   - A **prior pmf** π_{prior} which we think is the best estimate for π **before** we run the experiments.
5. **Output:** A **posterior pmf** π_{post} that we think is the best estimate for π **after** we observe the experiments.
6. **Method:**
   6.1 Compute the quantity
   $$K := \sum_{\theta \in \Omega} f_\theta(x_1) \ldots f_\theta(x_n) \pi_{prior}(\theta).$$
   6.2 Compute the posterior pmf π_{post} by
   $$\pi_{post}(\theta) = \frac{f_\theta(x_1) \ldots f_\theta(x_n)\, \pi_{prior}(\theta)}{K}.$$

Back to coin toss once again

Example 6.3

In the coin toss example. I pick one coin at random following an unknown pmf. In the language of Bayesian inference,

- X is a Bernoulli random variable with success probability Θ;
- Θ is randomly picked from the set $\{0.2, 0.5, 0.8\}$ following some unknown pmf π.
- Since we have no information regarding π, our best guess would be π_{prior} is the uniform distribution,

$$\pi_{prior}[\Theta = 0.2] \ = \ \pi_{prior}[\Theta = 0.5] \ = \ \pi_{prior}[\Theta = 0.8] = \frac{1}{3}.$$

- Now the chosen coin is flipped twice, and both outcomes are equal to head, so $x_1 = x_2 = 1$. **After** observing these experiments, we update our prediction on π by π_{post} using the given method.

Bayesian Inference(continuous)

Bayesian inference for the continuous case works the same way with discrete case, except that sum in the formula for K is replaced with integrals.

$$K := \int_{-\infty}^{\infty} f_\theta(x_1)\dots f_\theta(x_n)\pi_{prior}(\theta)\, d\theta.$$

Bayesian Inference(continuous)

Example 6.4

Let X be binomial distribution with parameters n (given) and θ (unknown),

$$f(\cdot|\theta) = f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad x = 0, 1, \ldots, n.$$

Suppose that π_{prior} is the beta pdf with parameter α, β (given),

$$\pi_{prior}(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad 0 < \theta < 1,$$

where Γ is the gamma function. Suppose that we have performed one experiment with outcome equal to x. Compute π_{post}.

$\Gamma(n) = (n-1)!$, n is an integer.

$$\begin{cases} \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta} \\ \text{Median} = \frac{\alpha - \frac{1}{3}}{\alpha+\beta-\frac{2}{3}} \end{cases}$$

$$K = \int_{-\infty}^{\infty} f_\theta(x)\, \pi_{prior}(\theta)\, d\theta$$

$$= \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1} \, d\theta$$

$$= \boxed{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}} \underbrace{\int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1} \, d\theta}$$

Beta $(x+\alpha,\ n-x+\beta)$

total probability $= 1$

$$\pi_{post}(\theta) = \frac{\cancel{\binom{n}{x}} \theta^x (1-\theta)^{n-x} \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)\Gamma(\beta)}} \theta^{\alpha-1} (1-\theta)^{\beta-1}}{\cancel{\binom{n}{x}} \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)\Gamma(\beta)}} \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}}$$

$$= C \cdot \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}$$

$$\propto \underbrace{\theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}}_{\text{density}}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)} \cdot \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}$$

$\sim$ Beta $(x+\alpha,\ n+\beta-x)$.

Bayesian Estimator

$\pi_{post}(\theta | x) \sim$ Distribution.

In the scenario of Bayes inference, the estimate for the unknown parameter Θ is not a *fixed number*, but a *random variable*. However, there are situations in real life where we are asked to give a fixed number $\hat{\theta}$ as our estimate.

$\min \mathbb{E}[(\Theta - b)^2] \Rightarrow b = \mathbb{E}[\Theta]$

Then **Bayesian estimator** $\hat{\theta}$ would depend on the penalty for errors created by incorrect guesses:

1. The loss function is $\mathbb{E}[(\Theta - \hat{\theta})^2]$, the **mean square error**. Then best guess $\hat{\theta}$ would be the mean of the posterior pdf.
2. The loss function is $\mathbb{E}[|\Theta - \hat{\theta}|]$, the **mean absolute error**. Then best guess $\hat{\theta}$ would be the median of the posterior pdf.

Bayesian Estimator

Example 6.5

Let X be the binomial random variable with parameters n and θ. Let π_{prior} be the beta pdf with parameters α and β. Suppose that we have one sample with value x. Compute the Bayesian estimator $\hat{\theta}$ that minimizes

- mean square error; $\hat{\theta} = \frac{\alpha+x}{(\alpha+x)+(n+\beta-x)}$
- mean absolute error, if $\alpha + x = n + \beta - x = 1$.

$$\text{Median} = \frac{\alpha+x-\frac{1}{3}}{(\alpha+x)+(n+\beta-x)-\frac{2}{3}} = \frac{1-\frac{1}{3}}{2-\frac{2}{3}} = \frac{\frac{2}{3}}{\frac{4}{3}} = \frac{1}{2}.$$

For an insurance Company, there are two risk classes.

prior 35% are smokers S

65% are non-smokers. S'

Annual claim count for each policy follows a Poisson distribution λ.

For smokers, λ has value of 3.

For non-smokers, λ has a value of 1. data/obs $N_1=3,\ N_2=0$

A randomly selected insured has 3 claims in Year 1 and no Claim in Year 2.

Calculate the expected number of claims by this insured in Year 3.

$$\mathbb{P}(\text{data}\,|\,S) = \mathbb{P}(N_1=3\,|\,S)\cdot\mathbb{P}(N_2=0\,|\,S) = \frac{e^{-3}3^3}{3!}\cdot\frac{e^{-3}3^0}{0!} = 0.011$$

$\mathbb{P}(S) = 0.35$ $\quad \mathbb{P}(S') = 0.65$ $\leftarrow$ priors.

$$\mathbb{P}(\text{data}\,|\,S') = \mathbb{P}(N_1=3\,|\,S')\,\mathbb{P}(N_2=0\,|\,S') = \frac{e^{-1}1^3}{3!}\ \frac{e^{-1}1^0}{0!} = 0.02256$$

$$K = \text{total prob} = \mathbb{P}(\text{data}) = 0.011\cdot 0.35 + 0.02256\cdot 0.65 = 0.01857$$

$$\text{posterior}: \quad \mathbb{P}(S\,|\,\text{data}) = \frac{0.011\cdot 0.35}{0.01857} = 0.21$$

$$\mathbb{P}(S'\,|\,\text{data}) = \frac{0.02256\cdot 0.65}{0.01857} = 0.79.$$

$$E[N_3\,|\,\text{data}] = E[N_3\,|\,S]\cdot\mathbb{P}(S\,|\,\text{data}) + E[N_3\,|\,S']\cdot\mathbb{P}(S'\,|\,\text{data})$$

$$= 3\cdot 0.21 + 1\cdot 0.79 = 1.4206$$