2. Posterior Statistics and Coin Flipping

Setup:

In this problem, we consider an application of posterior statistics. Suppose that we have five loaded coins; the probability of each landing heads is 0.2, 0.4, 0.4, 0.6, and 0.8. The coins are placed in a bag, and one coin is drawn completely at random. Let the probability that the coin drawn comes up heads on a given toss be λ_0. Our goal is to infer something about the probability that the coin picked will land heads, and we would extract information by flipping the coin n times and recording the outcomes of the tosses.

(a) Preliminary Ingredients

2.0/2 points (graded)

In a Bayesian model, given the description of the situation, what is the prior distribution of the parameter of interest λ, which we define as probability that the coin chosen lands heads?

Enter the prior probabilities for $\lambda = 0.2$, $\lambda = 0.4$, $\lambda = 0.6$, $\lambda = 0.8$, as a vector $(\mathbf{P}(\lambda = 0.2) \quad \mathbf{P}(\lambda = 0.4) \quad \mathbf{P}(\lambda = 0.6) \quad \mathbf{P}(\lambda = 0.8))$. For example, if $\mathbf{P}(\lambda = 0.2) = 0.5,\ \mathbf{P}(\lambda = 0.4) = 0.1,\ \mathbf{P}(\lambda = 0.6) = 0.1,\ \mathbf{P}(\lambda = 0.8) = 0.3,$ then enter **[0.5,0.1,0.1,0.3]**. Note the components are separated by commas, and the vector is enclosed by square brackets.

$(\mathbf{P}(\lambda = 0.2) \quad \mathbf{P}(\lambda = 0.4) \quad \mathbf{P}(\lambda = 0.6) \quad \mathbf{P}(\lambda = 0.8)) =$ [0.2,0.4,0.2,0.2] ✔ **Answer:** [0.2, 0.4, 0.2, 0.2]

Let the observations be $X_1, X_2, \cdots, X_n$, which are modelled as Bernoulli random variables indicating whether a head was tossed on each of the n tosses. Find a general expression for the likelihood function $L_n(X_1, X_2, \cdots, X_n | \lambda)$ in terms of λ, n, and $\sum_{i=1}^{n} X_i$.

(Enter **Sigma_i(X_i)** for $\sum_{i=1}^{n} X_i$. Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma_i(X_i)** by brackets.)

For this problem, write your answer in proportionality notation such that when $\lambda = 0.5$, the value of the likelihood function is 0.5^n regardless of the value of the X_i's.

$L_n(X_1, X_2, \cdots, X_n | \lambda) \propto$ lambda^(Sigma_i(X_i))*(1 ✔ **Answer:** lambda^(Sigma_i(X_i))*(1-lambda)^(n-(Sigma_i(X_i)))

STANDARD NOTATION

Solution:

Converting the coin frequencies into probabilities, we get that

$$\mathbf{P}(\lambda = 0.2) = \frac{1}{5},$$

$$\mathbf{P}(\lambda = 0.4) = \frac{2}{5},$$

$$\mathbf{P}(\lambda = 0.6) = \frac{1}{5},$$

and

$$\mathbf{P}(\lambda = 0.8) = \frac{1}{5}.$$

$L(X_i|\lambda)$ is a Bernoulli distribution with expected value λ, so its pmf is

$$L(X_i|\lambda) = \lambda^{X_i}(1-\lambda)^{1-X_i}.$$

Note that this form of the Bernoulli distribution pmf makes it especially easy to multiply; indeed, we could write

$$\begin{aligned} L_n(X_1,\ldots,X_n|\lambda) &= L(X_1|\lambda)\ldots L(X_n|\lambda) \\ &\propto \left(\lambda^{X_1}(1-\lambda)^{1-X_1}\right)\ldots\lambda^{X_n}(1-\lambda)^{1-X_n} \\ &\propto \left(\lambda^{X_1}\ldots\lambda^{X_n}\right)\left((1-\lambda)^{1-X_1}\ldots(1-\lambda)^{1-X_n}\right) \\ &\propto \lambda^{\sum X_i}(1-\lambda)^{n-\sum X_i}. \end{aligned}$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(b) Bayes' Theorem and the Posterior Distribution

4.0/4 points (graded)

Use Bayes' theorem to calculate the posterior distribution if $n=6$ and our observations are $X_1=0$, $X_2=X_3=1$, $X_4=0$, $X_5=X_6=1$. In this problem, the notational shorthand X_i refers to the vector of observations $X_1,\ldots,X_6$.

(Formatting a neat table in Excel may help in this.)

(Enter the posterior probabilities for $\lambda=0.2$, $\lambda=0.4$, $\lambda=0.6$, $\lambda=0.8$, as a vector $\left(\mathbf{P}(\lambda=0.2|X_i) \quad \mathbf{P}(\lambda=0.4|X_i) \quad \mathbf{P}(\lambda=0.6|X_i) \quad \mathbf{P}(\lambda=0.8|X_i)\right)$. For example if $\mathbf{P}(\lambda=0.2|X_i)=0.5,\ \mathbf{P}(\lambda=0.4|X_i)=0.1,\ \mathbf{P}(\lambda=0.6|X_i)=0.1,\ \mathbf{P}(\lambda=0.8|X_i)=0.2$ then enter **[0.5,0.1,0.1,0.2]**. Note the components are separated by commas, and the vector is enclosed by square brackets.)

$\left(\mathbf{P}(\lambda=0.2|X_i) \quad \mathbf{P}(\lambda=0.4|X_i) \quad \mathbf{P}(\lambda=0.6|X_i) \quad \mathbf{P}(\lambda=0.8|X_i)\right) =$ [0.0181,0.32579,0.366516,0.2896]

Answer: [0.018, 0.326, 0.367, 0.290]

From the posterior distribution, compute the posterior mean, mode, and median.

Posterior mean: 0.5855256 **Answer:** 0.59

Posterior mode: 0.6 **Answer:** 0.60

Posterior median: 0.6 **Answer:** 0.60

Solution:

By Bayes' theorem, $\pi(\lambda|X_i) \propto \pi(\lambda)L_n(X_i|\lambda)$. We recall the values $(0.2, 0.4, 0.2, 0.2)$ for $\pi(0.2)$, $\pi(0.4)$, $\pi(0.6)$, and $\pi(0.8)$. From the formula

$$L_n(X_i|\lambda) = \lambda^{\sum X_i}(1-\lambda)^{n-\sum X_i},$$

noting that $\sum X_i = 4$ and $n-\sum X_i = 2$, we can calculate

$$\begin{aligned} \left(L_n(X_i|0.2), L_n(X_i|0.4), L_n(X_i|0.6), L_n(X_i|0.8)\right) &= \left(0.2^4 0.8^2, 0.4^4 0.6^2, 0.6^4 0.4^2, 0.8^4 0.2^2\right) \\ &= (0.001024, 0.009216, 0.020736, 0.016384) \end{aligned}$$

Multiplying this with $\pi(\lambda)$ gives the un-normalized posterior distribution

$$\begin{aligned} \pi(X_i|\lambda) &\propto (0.2\cdot 0.001024, 0.4\cdot 0.009216, 0.2\cdot 0.020736, 0.2\cdot 0.016384) \\ &= (0.0002048, 0.0036864, 0.0041472, 0.0032768) \end{aligned}$$

Normalizing, this gives

$$\left(\pi(\lambda=0.2|X_i), \pi(\lambda=0.4|X_i), \pi(\lambda=0.6|X_i), \pi(\lambda=0.8|X_i)\right) = (0.0181, 0.3258, 0.3665, 0.2896).$$

From this, we could compute the posterior mean to be

$$\sum_{j\in\{0.2,0.4,0.6,0.8\}} \pi(\lambda=j|X_i)\lambda = (0.0181)(0.2)+(0.3258)(0.4)+(0.3665)(0.6)+(0.2896)(0.8) = 0.5855.$$

注意是值（硬币的概率）乘以权重（硬币不同概率的后验概率）

To compute the median, notice that

$$\pi\left(\lambda=0.2|X_i\right)+\pi\left(\lambda=0.4|X_i\right)=0.0181+0.3258=0.4439<0.5000$$

and that

$$\pi\left(\lambda=0.2|X_i\right)+\pi\left(\lambda=0.4|X_i\right)+\pi\left(\lambda=0.6|X_i\right)=0.0181+0.3258+0.3665=0.8104>0.5000,$$

meaning that the median is at 0.6. From the posterior distribution, we could also see that the λ with the highest mass is $\lambda=0.6$ with frequency 0.3665, so the mode is 0.6.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(c) Choosing Between Estimators

2/2 points (graded)
The estimator we choose depends on the situation. If we are only allowed to name one estimate for the parameter λ_0 as our "best guess" incorporating our observations, which among the three statistics named in the previous problem may be used for sure, regardless of our observations (i.e. it is always uniquely defined)? You should pick the estimator(s) that would always produce a unique value.

(Choose all that apply.)

- [x] Posterior mean ✔
- [ ] Posterior median
- [ ] Posterior mode

✔

Consider the general case with n observations. Suppose that the coin we picked is known to a certain observer, and we are to state an estimate λ_0 of the probability that the coin we picked lands heads at a given toss. The observer will reward a certain amount of money depending on both our estimate and the actual value, which will be one of the three systems below.

1. We will be rewarded $\$100$ for the correct value (i.e. $\lambda_0=\lambda^*$) and $\$0$ otherwise.
2. We will be rewarded $\$\left(100-100|\lambda_0-\lambda^*|\right)$.
3. We will be rewarded $\$\left(100-\frac{(\lambda_0-\lambda^*)^2}{100}\right)$.

For each of the three systems, identify the most appropriate estimator among the posterior mean, median, or mode. Select the choice corresponding to your answer in order (first, second, then third item).

- ( ) mean, median, mode
- ( ) mean, mode, median
- ( ) median, mean, mode
- ( ) median, mode, mean
- ( ) mode, mean, median
- (•) mode, median, mean ✔

Solution:

Unique Value-Producing Estimators

- The posterior mean will always produce a unique estimate for λ_0 as it's defined as the weighted average of the distribution, and has thus has an arithmetic formula.
- The posterior median is not necessarily uniquely defined because it's possible say for exactly half the weight in the posterior distribution to lie on the left half (0.2, 0.4) and half on the right half (0.6, 0.8). In this case, the median can be any real number from $[0.4, 0.6]$, so it's not uniquely defined.

- The posterior mode is not necessarily uniquely defined because it's possible for the posterior distribution to be bimodal.

Estimators and Reward Functions This question is intended as a review of some concepts from M-estimation.

- In this case, we wish to optimize for probability that we pick the correct value. As we have a discrete set of observations, we have to pick the coin with the largest probability of being the correct coin. This is thus the **posterior mode**.

- In this case, we are optimizing for the minimum expected absolute distance. We have shown here that the median minimizes the expected absolute distance, and because our distribution is based on the posterior distribution, we have to pick the **posterior median**. Intuitively, one could reason that we could pair up equivalent areas starting from the left and right of the posterior distribution. Then we have the property that as long as we're inside the interval formed by these two points, the total absolute distance is equal, and strictly greater once we're outside the interval. It is then optimal to pick the median so that we're never outside the interval.

- In this case, we are minimizing the minimum expected square distance. Again, we have shown here that the mean minimizes the expected square distance to any point in a distribution. As our distribution in this situation is the posterior distribution, we have to pick the **posterior mean**. Indeed, we are effectively minimizing $\mathbb{E}\left[(\lambda_0 - \lambda^*)^2\right]$, which can be written as a quadratic in λ_0:

$$\lambda_0^2 - 2\mathbb{E}\left[\lambda^*\right]\lambda_0 + \mathbb{E}\left[(\lambda^*)^2\right].$$

  The coefficients of this quadratic are 1, $-2\mathbb{E}\left[\lambda^*\right]$ and $\mathbb{E}[\lambda^*)^2]$, so its minimum is at $\mathbb{E}\left[\lambda^*\right]$, indeed the mean.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Homework 9: Bayesian Statistics / 2. Posterior Statistics and Coin Flipping