WEEK2.8 BAYES

Created	@April 9, 2024 3:32 PM
Class	ADS
Reviewed	☐

Q1 Is the Guinness factory adding enough barley?

The Guinness beer factory requires 50 g of barley per pint of beer. They examined 50 pints and found anaverage barley content of 46 g per pint.

1. Use my priors as described below to determine the posterior probability that enough barley is being added to each pint.
   P(H1) = P(enough barley) = 0.5
   P(H2) = P(not enough barley) = 0.5
   P(DATA|H1) = P(mean barley content 46g|enough barley) = 0.7
   P(DATA|H2) = P(mean barley content 46g|not enough barley) = 0.4
   What is the Bayes Factor for these two hypotheses?
2. Lets explore the probability of seeing the data given each of the two hypotheses. The numbers given here are estimates based on experience. To see how they influence the Bayes factor, calculate the Bayes Factor for each combination of P(DATA|H1)and P(DATA|H2) from 0 to 1 in steps of 0.1 Please plot these on a graph where these two probabilities are the x and y axes.
3. Now, I think that my staff are stealing barley to make their own moonshine Guiness at home. I update my first prior to P(A) = P(enough barley) = 0.2. How does this affect the Bayes Factor? How does it affect the posteriors?

ANSWER

1. **后验概率与贝叶斯因子**：

   首先，我们需要计算后验概率。后验概率是根据贝叶斯定理计算的，公式为：

   $P(H_i|DATA) = \frac{P(DATA|H_i)P(H_i)}{P(DATA)}$

其中，P(DATA)$是观测到数据的概率，可以通过求和所有假设下的数据概率和先验概率的乘积来计算，即：

$$P(DATA) = P(DATA|H1)P(H1) + P(DATA|H2)P(H2) = 0.7 \times 0.5 + 0.4 \times 0.5 = 0.55$$

然后，我们可以计算后验概率：

$$P(H1|DATA) = \frac{P(DATA|H1)P(H1)}{P(DATA)} = \frac{0.7 \times 0.5}{0.55} = 0.6364$$

$$P(H2|DATA) = \frac{P(DATA|H2)P(H2)}{P(DATA)} = \frac{0.4 \times 0.5}{0.55} = 0.3636$$

贝叶斯因子是比较两个假设下的数据概率的一种方法，公式为：

$$BF = \frac{P(DATA|H1)}{P(DATA|H2)} = \frac{0.7}{0.4} = 1.75$$

如果贝叶斯因子大于1，那么数据更倾向于支持假设 $H1$；如果贝叶斯因子小于1，那么数据更倾向于支持假设 H2。在这个例子中，贝叶斯因子大于1，所以数据更倾向于支持有足够大麦的假设。

2. **贝叶斯因子的图形表示**：

   这部分需要使用编程工具来计算和绘制图形，例如Python或R。你需要在一个循环中改变 $P(DATA|H1)$ 和 $P(DATA|H2)$ 的值，然后计算每个组合下的贝叶斯因子，最后将这些贝叶斯因子作为颜色或等高线绘制在二维平面上。

3. **更新先验概率**：

   如果你更新了假设 $H1$ 的先验概率，那么后验概率和贝叶斯因子都会改变。新的后验概率和贝叶斯因子可以通过上述公式重新计算。具体来说，$P(H1)$ 变为 0.2，$P(H2)$ 因此变为 0.8（因为 $P(H1) + P(H2) = 1$）。

   新的 $P(DATA)$ 为：

   $$P(DATA) = P(DATA|H1)P(H1) + P(DATA|H2)P(H2) = 0.7 \times 0.2 + 0.4 \times 0.8 = 0.46$$

   新的后验概率为：

   $$P(H1|DATA) = \frac{P(DATA|H1)P(H1)}{P(DATA)} = \frac{0.7 \times 0.2}{0.46} = 0.3043$$

   $$P(H2|DATA) = \frac{P(DATA|H2)P(H2)}{P(DATA)} = \frac{0.4 \times 0.8}{0.46} = 0.6957$$

   新的贝叶斯因子为：

   $$BF = \frac{P(DATA|H1)}{P(DATA|H2)} = \frac{0.7}{0.4} = 1.75$$

   注意，贝叶斯因子并没有改变，因为它只依赖于数据和假设，而与先验概率无关。然而，后验概率发生了显著变化，现在数据更倾向于支持没有足够大麦的假

设。

Q2 A dice game

You play a dice-based game with a friend online, using six-sided dice. In this game, rolling a six is especially lucky and wins you many points.
Since you are playing via video link, your friend uses their own die. You can see that it is six-sided, but you cannot see the numbers clearly.
Over the course of the game, your friend rolls the die 20 times, and 7 out of those 20 rolls are sixes.
Do you think your friend is playing fair? How many sixes do you believe their die has? 1 (like a normal die)? Or more (the die is unfair)?
How? Before you start coding, think about this for a moment. The main idea is you have 6 possible hypotheses, one for each number of sixes. Is it really 6 though? Also, how do you compare these hypothesis? Do you have to compare every hypothesis with every other?

ANSWER

```
# Number of trials and successes
n_trials <- 20
n_successes <- 7
```

这两行代码定义了试验的次数和成功的次数。在这个例子中，试验的次数是你的朋友掷骰子的次数（20次），成功的次数是掷出6的次数（7次）。

```
# Prior probabilities
priors <- rep(1/6, 6)
```

这行代码定义了每个假设的先验概率。在这个例子中，我们假设每个假设在开始时都是等可能的，所以先验概率都是1/6。

```
# Likelihoods
likelihoods <- dbinom(n_successes, size = n_trials, prob =
(1:6)/6)
```

这行代码计算了每个假设的似然性。在这个例子中，似然性是在给定的试验次数（20次掷骰子）和成功概率（掷出6的概率）下观察到给定的成功次数（7次掷出6）的概率。这是一个二项分布问题，可以使用R的`dbinom`函数来计算。

```
# Unnormalized posteriors
unnormalized_posteriors <- priors * likelihoods
```

这行代码计算了每个假设的未归一化的后验概率。在贝叶斯推断中，后验概率是似然性乘以先验概率。然而，这样计算出来的后验概率可能不会加起来等于1，所以我们称它们为未归一化的后验概率。

```
# Normalized posteriors
posteriors <- unnormalized_posteriors / sum(unnormalized_posteriors)
```

这行代码计算了每个假设的归一化的后验概率。我们需要将每个假设的未归一化的后验概率除以所有假设的未归一化的后验概率的总和，以确保后验概率的总和为1。

```
# Print the posteriors
print(posteriors)
```

这行代码打印出每个假设的后验概率。你可以查看哪个假设的后验概率最高，那就是最可能的假设。

总的来说，这段代码使用贝叶斯推断来解答关于骰子公平性的问题。通过计算每个假设的似然性和先验概率，并使用贝叶斯定理，我们可以得到每个假设的后验概率。最后，我们可以比较后验概率，找出最可能的假设。