TA Lecture 02 - Conditional Probability

Oct 24-26

School of Information Science and Technology,
ShanghaiTech University

Outline

Main Contents Recap

HW Problems

More Exercices

Conditioning

- New data & information may affect our uncertainties
- Conditional probability: how to update our belief?
- All probabilities are conditional! (explicit/implicit background info or assumption)

Conditioning

- Conditioning :a powerful problem-solving strategy
- Reducing a complicated probability problem to a bunch of simpler conditional probability problems
- First-step analysis: obtain recursive solution to multi-stage problems
- **Conditioning is the soul of statistics!**

Conditional Probability

Definition

If A and B are events with $P(B) > 0$, then the *conditional probability* of A given B, denoted by $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

- $P(A)$: *prior probability* of A.
- $P(A|B)$: *posterior probability* of A.

Bayes' Rule

Theorem

For any events A and B with positive probabilities,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

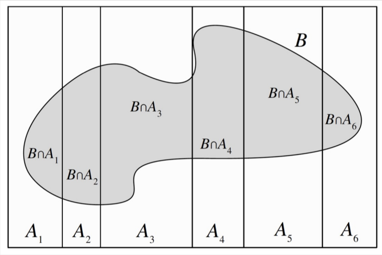
Bayes' Rule with Extra Conditioning

Theorem

Provided that $P(A \cap E) > 0$ *and* $P(B \cap E) > 0$, *we have*

$$P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}.$$

LOTP



Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

LOTP with Extra Conditioning

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i \cap E) > 0$ for all i. Then

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E).$$

Independence

Definition

Events A and B are *independent* if

$$P(A \cap B) = P(A)P(B).$$

If $P(A) > 0$ and $P(B) > 0$, then this is equivalent to

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Independence for Multiple Events

Definition

Events A, B and C are *independent* if all of the following equations hold:

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B) \\
P(A \cap C) &= P(A)P(C) \\
P(B \cap C) &= P(B)P(C) \\
P(A \cap B \cap C) &= P(A)P(B)P(C).
\end{aligned}
$$

Conditional Independence

Definition

Events A and B are said to be *conditionally independent* given E if:

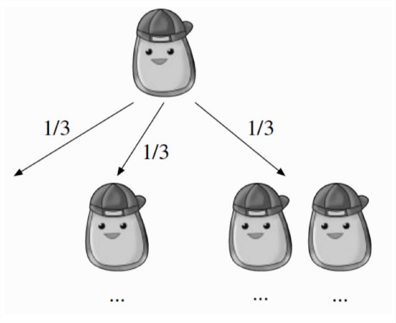
$$P(A \cap B|E) = P(A|E)P(B|E).$$

Independence vs. Conditional Independence

- We choose either a fair coin or a biased coin (w.p. 3/4 of landing Heads).
- But we do not know which one we have chosen and we flip it twice.
- Event F: “chosen the fair coin”
- Event A_1: “the first coin tosses landing Heads”
- Event A_2: “the second coin tosses landing Heads”

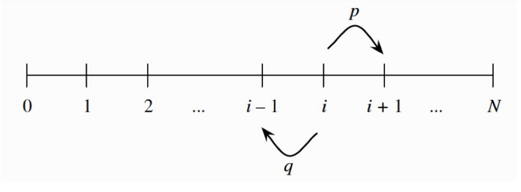
Probability Model: Branching Process

A single amoeba, Bobo, lives in a pond. After one minute Bobo will either die, split into two amoebas, or stay the same, with equal probability, and in subsequent minutes all living amoebas will behave the same way, independently. What is the probability that the amoeba population will eventually die out?

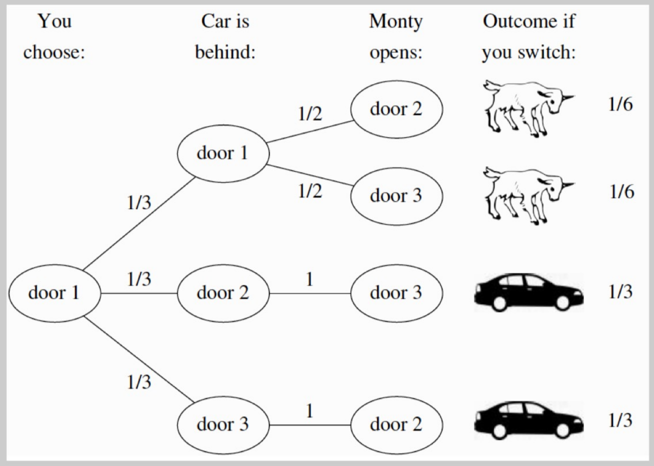


Probability Model: Gambler's Ruin

Two gamblers, A and B, make a sequence of dollar 1 bets. In each bet, gambler A has probability p of winning, and gambler B has probability $q = 1 - p$ of winning. Gambler A starts with i dollars and gambler B starts with $N - i$ dollars; the total wealth between the two remains constant since every time A loses a dollar, the dollar goes to B, and vice versa. The game ends when either A or B is ruined (run out of money). What is the probability that A wins the game (walking away with all the money)?



Probability Model: Monty Hall



Outline

Main Contents Recap

HW Problems

More Exercices

Problem 1

Alice is trying to communicate with Bob, by sending a message (encoded in binary) across a channel.

(a) Suppose for this part that she sends only one bit (a 0 or 1), with equal probabilities. If she sends a 0, there is a 5% chance of an error occurring, resulting in Bob receiving a 1; if she sends a 1, there is a 10% chance of an error occurring, resulting in Bob receiving a 0. Given that Bob receives a 1, what is the probability that Alice actually sent a 1?

Problem 1 Solution

(b) To reduce the chance of miscommunication, Alice and Bob decide to use a repetition code. Again Alice wants to convey a 0 or a 1, but this time she repeats it two more times, so that she sends 000 to convey 0 and 111 to convey 1. Bob will decode the message by going with what the majority of the bits were. Assume that the error probabilities are as in (a), with error events for different bits independent of each other. Given that Bob receives 110, what is the probability that Alice intended to convey a 1?

Problem 1 Solution

Problem 2

Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

(a) Assume for this part that the test results are conditionally independent given Freds disease status. Let $a = P(T_j|D)$ and $b = P(T_j|D^c)$, where a and b dont depend on the jth test. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the n tests.

Problem 2 Solution

Problem 2 Continued

(b) Suppose that Fred tests positive on all n tests. However, some people have a certain gene that makes them always test positive. Let G be the event that Fred has the gene. Assume that $P(G) = 1/2$ and that D and G are independent. If Fred does not have the gene, then the test results are conditionally independent given his disease status. Let $a_0 = P(T_j|D, G^c)$ and $b_0 = P(T_j|D^c, G^c)$, where a_0 and b_0 dont depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the tests.

Problem 2 Solution

Problem 3

We want to design a spam filter for email. As described in Exercise 1, a major strategy is to find phrases that are much more likely to appear in a spam email than in a non-spam email. In that exercise, we only consider one such phrase: “free money”. More realistically, suppose that we have created a list of 100 words or phrases that are much more likely to be used in spam than in non-spam. Let W_j be the event that an email contains the jth word or phrase on the list.

Problem 3 Continued

Let

$$p = P(\text{spam}), p_j = P(W_j|\text{spam}), r_j = P(W_j|\text{not spam})$$

where "spam" is shorthand for the event that the email is spam. Assume that $W_1, ..., W_{100}$ are conditionally independent given M (the event that the email is spam), and also conditionally independent given M^c (the event that the email is non-spam). A method for classifying emails (or other objects) based on this kind of assumption is called a *naive Bayes classifier*. (Here "naive" refers to the fact that the conditional independence is a strong assumption, not to Bayes being naive. The assumption may or may not be realistic, but naive Bayes classifiers sometimes work well in practice even if the assumption is not realistic.)

Problem 3 Continued

Under this assumption we know, for example, that

$$P(W_1, W_2, W_3^c, W_4^c, \ldots, W_{100}^c|\text{spam}) = p_1 p_2 (1-p_3)(1-p_4)\cdots(1-p_{100}).$$

Without the naive Bayes assumption, there would be vastly more statistical and computational difficulties since we would need to consider $2^{100} \approx 1.3 \times 10^{30}$ events of the form $A_1 \cap A_2 ... \cap A_{100}$ with each A_j equal to either W_j or W_j^c. A new email has just arrived, and it includes the 23rd, 64th, and 65th words or phrases on the list (but not the other 97).

Problem 3 Continued

So we want to compute

$$P(\text{spam}|W_1^c, \ldots, W_{22}^c, W_{23}, W_{24}^c, \ldots, W_{63}^c, W_{64}, W_{65}, W_{66}^c, \ldots, W_{100}^c).$$

Note that we need to condition on *all* the evidence, not just the fact that $W_{23} \cap W_{64} \cap W_{65}$ occurred. Find the conditional probability that the new email is spam (in terms of p and the p_j and r_j).

Problem 3 Solution

Problem 4

In Monty Hall problem, now suppose the car is not placed randomly with equal probability behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly with equal probability among his options in those cases where your initial choice is correct. What strategy should you follow?

Problem 4 Solution

Problem 5

Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $1/2 \leq p \leq 1$. To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is door 1.

(a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).

Problem 5 Solution

Problem 5 Continued

(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 2.

(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 3.

Problem 5 Solution

Problem 6

A/B testing is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements.

Problem 6 Continued

As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as “success” (e.g., the user made a purchase) or “failure”. The probability that the n-th user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a two-armed bandit. Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a “stay-with-a-winner” procedure: (i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities. (ii) If the trial for the n-th user is a success, stay with the same treatment for the $(n + 1)$-st user; otherwise, switch to the other treatment for the $(n + 1)$-st user.

Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the n-th trial and a_n be the probability that Treatment A is assigned on the n-th trial (using the above algorithm).

(a) Show that

$$p_n = (a - b)a_n + b, \; a_{n+1} = (a + b - 1)a_n + 1 - b.$$

Problem 6 Solution

Problem 6 Continued

(b) Use the results from (a) to show that p_{n+1} satisfies the following recursive equation:

$$p_{n+1} = (a + b - 1)p_n + a + b - 2ab.$$

(b) Use the result from (b) to find the long-run probability of success for this algorithm, $\lim_{n\to+\infty} p_n$, assuming that this limit exists.

Problem 6 Solution

Problem 7

In a game hosted by Monty, you are presented with n identical doors, with $n \geq 3$. You select one, but do not open it. Monty now opens a door he knows to conceal a goat, and gives you the option of switching doors. After making your choice, Monty reveals another goat. He again gives you the option of switching. This process continues until only two doors remain (your current choice, and one other unopened door). You make your final choice, and receive whatever is behind your door. We assume throughout that Monty always chooses randomly from among the goat-concealing doors when more than one such door remains in play.

Problem 7 Continued

(a) What strategy maximizes your chances of success?

(b) Under such a strategy, what is your winning probability?

Problem 7 Solution

Outline

Main Contents Recap

HW Problems

More Exercices

BH CH2 #38(b)

Consider the following n-door version of the Monty Hall problem. There are n doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens m goat doors, and offers you the option of switching to any of the remaining $n - m - 1$ doors.

Assume that Monty Hall knows which door has the car, will always open m goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch?

Solution