

MIT OpenCourseWare
<http://ocw.mit.edu>

14.30 Introduction to Statistical Methods in Economics
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

14.30 Introduction to Statistical Methods in Economics

Lecture Notes 4

Konrad Menzel

February 12, 2009

1 Bayes Theorem

In the last lecture, we introduced conditional probabilities, and we saw the Law of Total Probability as a way of relating the unconditional probability $P(A)$ of an event A to the conditional probabilities $P(A|B_i)$. Another important relationship between conditional probabilities is Bayes Law which relates the conditional probability $P(A|B)$ to the conditional probability $P(B|A)$, i.e. how we can revert the order of conditioning. This result plays an important role in many areas of statistics and probability, most importantly in situations in which we "learn" about the "state of the world" A from observing the "data" B .

Example 1 *The ancient Greeks (who apparently didn't know much statistics yet) noticed that each time after a ship had sunk, all surviving seamen reported having prayed to Poseidon, the Greek god of the sea. From this observation, they inferred that they were in fact saved from drowning because they had prayed. This example was actually brought up by the English philosopher Francis Bacon in the 16th century. In statistical terms, let's define the events $A = \text{"survives"}$ and $B = \text{"prayed"}$, so that the question becomes whether praying increases the odds of survival, i.e. whether $P(A|B) > P(A) \equiv p$, say. The observation that all surviving seaman had been praying translates to $P(B|A) = 1$. Is that information actually sufficient to answer the question whether praying strictly increases the chances of survival? How do we use the information on $P(B|A)$ to learn about $P(A|B)$?*

From the definition of conditional probabilities, we obtain

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

Rearranging the second equality, we get

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

We've also seen that we can partition the event

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

so that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

We can generalize this to any partition of S as summarized in the following theorem:

Theorem 1 (Bayes Theorem) If A_1, A_2, \dots is a partition of S , for any event B with $P(B) > 0$ we can write

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j \geq 1} P(B|A_j)P(A_j)}$$

- $P(A_i)$ is the *prior* probability of an event A_i (i.e. probability before experiment is run)
- $P(A_i|B)$ is the *posterior* probability of A_i (i.e. the probability after we ran the experiment and got information B - as obtained from Bayes theorem)

An entire statistical theory of optimal decisions is built on this simple idea: Bayesian Decision Theory.

Example 2 For the previous example of seamen surviving the sinking of a ship, we were able to observe $P(B|A) = 1$, and the (unconditional) survival rate of seamen, $P(A)$. However, we can also see that we don't have sufficient information to answer the question whether praying strictly increases the chances of survival since we couldn't observe $P(B|A^C)$, the fraction of seamen who prayed among those who were going to drown. It is probably safe to assume that those, fearing for their lives, all of them prayed as well (implying $P(B|A^C) = 1$), so that

$$P(A|B) = \frac{1 \cdot p}{1 \cdot p + 1 \cdot (1 - p)} = p = P(A)$$

In a sense, the reasoning of the ancient Greeks is an instance of "survivor bias" (well, in a very literal sense): Bayes theorem shows us that if we can only observe the survivors, we can't make a judgement about why they survived unless we know more about the subpopulation which did not survive.

Example 3 An important application of Bayes rule is how we should interpret a medical test. Suppose a doctor tests a patient for a very nasty disease, and we'll call the event that the patient in fact has the disease A . The test can either give a positive result - we'll call this event B - or a negative result, corresponding to event B^C . The test is not fully reliable in the sense that we can't determine for sure whether the patient has the disease, but the probabilities of a positive test result is

$$P(B|A) = 99\%, \quad P(B|A^C) = 5\%$$

Finally, we know that the disease is relatively rare and affects 0.5% of the population which shares the patient's age, sex, and other characteristics. Let's say the test gives a positive result. What is the (conditional) probability that the patient does in fact have the disease?

Bayes rule gives that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.05 \cdot 0.995} \approx 0.0905$$

Since the overall prevalence of the disease, $P(A)$ is relatively low, even a positive test result gives only relatively weak evidence for disease.

Example 4 Romeo and Juliet have been dating for some time, and come Valentine's Day (as a reminder: that's Saturday), Romeo can give Juliet either jewelery, J , or a serenade, S . Juliet wants jewelery, and if Romeo really loved her, he would read her wishes from her eyes, and besides, she had told Romeo about the jewelery two weeks earlier, during the final half hour of the Superbowl. Juliet also has first doubts that Romeo still loves her, an event we'll call L . To be specific

$$P(L) = 0.95$$

Juliet also knows that if Romeo loved her, he would give her jewelery with probability $P(J|L) = 0.80$, or a serenade with probability

$$P(S|L) = 0.20$$

(this is actually only what Juliet thinks, keeping in mind that Romeo also loves football). If he doesn't love her anymore, he has no idea what Juliet wants, and he'll give her a serenade (or, more realistically, the roses she wanted last year, or forget about Valentine's Day altogether) with probability

$$P(S|L^C) = 0.80$$

(note that, though giving a serenade is still very embarrassing for Romeo, it's also much cheaper). It turns out that Romeo ends up giving Juliet a serenade. Should she dump him right away? By Bayes' theorem, Juliet's posterior beliefs about Romeo's intentions are

$$P(L|S) = \frac{P(S|L)P(L)}{P(S|L)P(L) + P(S|L^C)P(L^C)} = \frac{(1 - 0.8)0.95}{(1 - 0.8)0.95 + 0.8(1 - 0.95)} = \frac{\frac{2}{10} \cdot \frac{19}{20}}{\frac{2}{10} \cdot \frac{19}{20} + \frac{8}{10} \cdot \frac{1}{20}} = \frac{38}{46} \approx 0.826$$

and we'll let Juliet decide on her own whether this is still good enough for her.

In real-life situations, most people aren't very good at this type of judgments and tend to overrate the reliability of a test like the ones from the last two examples - in the cognitive psychology literature, this phenomenon is known as *base-rate neglect*, where in our example "base-rate" refers to the proportions $P(A)$ and $P(A^C)$ of infected and healthy individuals, or the prior probabilities $P(L)$ and $P(L^C)$ of Romeo loving or not loving Juliet, respectively. If these probabilities are very different, biases in intuitive reasoning can be quite severe.

Example 5 The "Monty Hall paradox":¹ There used to be a TV show in which a candidate was asked to choose among three doors A, B, C . Behind one of the doors, there was a prize (say, a brand-new washing machine), and behind each of the other two there was a goat. If the candidate picked the door with the prize, he could keep it, if he picked a door with a goat, he wouldn't win anything. In order to make the game a little more interesting, after the candidate made an initial choice, the host of the show would always open one of the two remaining doors with a goat behind it. Now the candidate was allowed to switch to the other remaining door if he wanted. Would it be a good idea to switch?

Without loss of generality, assume I picked door A initially. The unconditional probability of the prize being behind door A is $\frac{1}{3}$. If the prize was in fact behind door a , the host would open door b or door c , both of which have a goat behind them, with equal probability. If the initial guess was wrong, there is only one door left which was neither chosen by the candidate nor contains the prize. Therefore, given I initially picked A and the host then opened C ,

$$P(\text{prize behind } A | C \text{ opened}) = \frac{P(C \text{ opened} | \text{prize behind } A)P(\text{prize behind } A)}{P(C \text{ opened})} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

On the other hand

$$P(\text{prize behind } B | C \text{ opened}) = \frac{P(C \text{ opened} | \text{prize behind } B)P(\text{prize behind } B)}{P(C \text{ opened})} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Therefore, I could increase my chances of winning the prize if I switch doors. Intuitively, the newly opened door does not convey any information about the likelihood of the prize being

¹You can read up on the debate on <http://people.csail.mit.edu/carroll/probSem/Documents/Monty-NYTimes.pdf>

behind door A, since the host would not have opened it in any case - and in fact, given that we chose A, the events "A contains the prize" and "C is opened" are independent. However, the fact that he did not open B could arise from two scenarios: (1) the prize was behind A, and the host just chose to open C at random, or (2) the prize was behind door B, so the host opened door C only because he didn't have a choice. Therefore, being able to rule out C only "benefits" event B.

2 Recap Part 1: Probability

Before we move on to the second chapter of the lecture let's just summarize what we have done so far, and what you should eventually should feel familiar and comfortable with:

2.1 Counting Rules

- drawing n out of N with replacement: N^n possibilities
- drawing n out of N without replacement: $\frac{N!}{(N-n)!}$ possibilities
- permutations: $N!$ possibilities
- combinations of n out of N : $\binom{N}{n}$ possibilities

2.2 Probabilities

- independence: $P(AB) = P(A)P(B)$
- conditional probability $P(A|B) = \frac{P(AB)}{P(B)}$ if $P(B) > 0$.
- $P(A|B) = P(A)$ if and only if A and B are independent
- Law of Total Probability: for a partition B_1, \dots, B_n of S , where $P(B_i) > 0$

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

- Bayes Theorem:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

There are also a few things we saw about probabilistic reasoning in general:

- use of set manipulations to reformulate event of interest into something for which it's easier to calculate probabilities (e.g. complement, partitions etc.)
- the importance of base rates in converting conditional into marginal/unconditional probabilities (e.g. in Bayes theorem or composition effects in the heart surgeons example)
- can sometimes make dependent events A and B independent by conditioning on another event C (or make independent events dependent).

3 Random Variables

Now let's move on to the second big topic of this class, random variables.

Example 6 *Flipping a coin, version I: We could define a variable X which takes the value 1 if the coin comes up Heads H and 0 if we toss Tails T . The sample space for this random experiment is $S = \{H, T\}$, and the range of the random variable is $\{0, 1\}$.*

Definition 1 *A real-valued random variable X is any function*

$$X : \begin{cases} S \rightarrow \mathbb{R} \\ s \mapsto X(s) \end{cases}$$

which maps the outcomes of an experiment to the real numbers.

As a historical aside, when the idea of random variables was developed around 1800, there was no role for "genuine" randomness in the minds of mathematicians and other scientists. Rather, chance was seen as a consequence of us not having full knowledge about all parameters of a situation we are analyzing, and our inability to apply the (supposedly fully deterministic) laws of nature to predict the outcome of an experiment. A being able to do all this is known as the "Laplacean demon", described by the famous mathematician Pierre Simon de Laplace as follows:

An intellect which at any given moment knew all the forces that animate Nature and the mutual positions of the beings that comprise it, if this intellect were vast enough to submit its data to analysis, could condense into a single formula the movement of the great bodies of the universe and that of the lightest atom: for such an intellect nothing could be uncertain, and the future just like the past would be present before its eyes.²

This view of the world doesn't quite hold up to subsequent developments in physics (e.g. genuine indeterminacy in quantum physics) or computational theory (e.g. Gödel's theorem: the intellect would have to be more complex than itself since its predictions are part of the universe it's trying to predict), but it's still what underlies our basic concept of probabilities: randomness in the world around us primarily reflects our lack of information about it.

Example 7 *As an illustration, here is version II of the coin flip: in order to illustrate Laplace's idea, we could think of a more complicated way of defining the sample space than in the first go above: in classical mechanics, we can (at least in principle) give a full description of the state of the coin (a rigid body) at any point in time, and then use the laws of classical mechanics to predict its full trajectory, and in particular whether it is going to end up with heads (H) or tails (T) on top. More specifically, we could describe the sample space as the state of the mechanical system at the point in time the coin is released into the air. A full (though somewhat idealized) description of the state of the system is given by (1) the position, and (2) the velocity of the center of mass of the coin together with (3) its orientation, and (4) its angular momentum at a given time t_0 - each of these has 3 coordinates, so $S = \mathbb{R}^{12}$. Each point $s \in S$ belongs to one of the two events $\{H, T\}$ deterministically. If we assign values $X = 1$ to the event $H \subset S$ that heads are up, and $X = 0$ for tails T , this mapping is the random variable*

$$X : \mathbb{R}^{12} \rightarrow \{0, 1\}, \quad \text{given by } X : \begin{cases} s \mapsto 1 & \text{if } s \in H \\ s \mapsto 0 & \text{otherwise, i.e. if } s \in T \end{cases}$$

Since the problem is - almost literally - very knife-edged, the outcome is highly sensitive to tiny changes

²Laplace, P. (1814): A Philosophical Essay on Probabilities

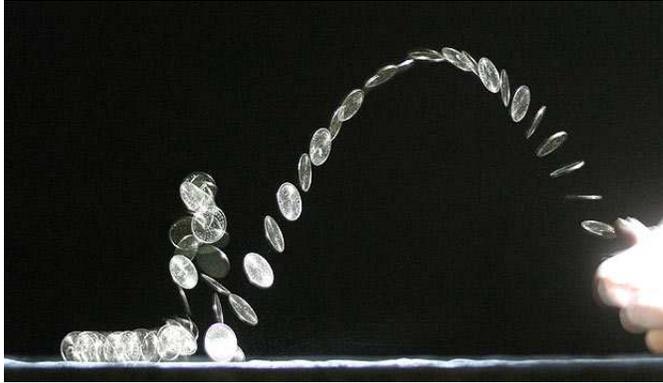


Figure 1: Stroboscopic image of a coin flip (Courtesy of Andrew Davidhazy and the School of Photo Arts and Sciences of Rochester Institute of Technology. Used with permission. ©Andrew Davidhazy, 2007.)

in the initial state $s \in S$ (not even taking into account external influences like, say, the gravitation of a car driving by), and no matter how we flip the coin, it is totally impossible to control the initial position, velocity, etc. precisely enough to produce the desired outcome with certainty. Also, it will typically be impossible to solve the differential equations describing the system with sufficient precision. Therefore, we can only give probabilities for being in parts of S , which again maps to probabilities for the outcomes H and T . So even though there need not be any "genuine" randomness in the situation, this is how it plays out for us in practice.

While this definition of the random variable brings out more the philosophical point how we think about random variables and probabilities, it is clearly not very operational, and for practical purposes, we'd rather stick to the first way of describing the problem.

Remark 1 The probability function on the sample space S induces a probability distribution for X via

$$P(X \in A) = P(\{s \in S : X(s) \in A\})$$

for any event $A \subset \mathbb{R}$ in the range of X .

Even though formally, X is a function from the sample space into the real numbers, we often treat it as a variable, i.e. we say that it can "take on" various values with the corresponding probabilities without specifying its argument. In other words, for most applications, we will only specify $P(X \in A)$ without any reference to the underlying sample space S and the probability on S . E.g. in the coin flip example - as described above - we won't try figure out the exact relationship between coordinates in S (initial position, velocity, orientation etc. of the coin) and outcomes (which is numerically impossible) and a probability distribution over these coordinates, but we just need to know that $P(X = 1) = P(X = 0) = \frac{1}{2}$.

Example 8 If we toss 10 coins independently of another, we could define a random variable $X = (\text{Total Number of Tails})$. We'll analyze the distribution for this type of random variable in more detail below.

Example 9 We might be interested in the outcome of an election. Say there are 100 million voters and 2 candidates. Each voter can only vote for either of the candidates, so there are $2^{100,000,000}$ distinct outcomes in terms of which voter voted for which candidate. We could now define the random variable X corresponding to the total number of votes for candidate A (and for elections, that's typically all we care about). For each value for the number of votes, there is a corresponding number of basic outcomes,

e.g. there is only one way candidate A can get all the votes. We can formulate the probability of each outcome in terms of simple probabilities, and from there we can aggregate over equivalent outcomes to obtain the probability for a given total number of votes.

Remark 2 *Not all random events have a numerical characteristic associated with them in which we are interested (e.g. if the event is "it will rain tomorrow", we might not care how much). Then we don't have to bother with random variables, but can just deal with events as before. Alternatively, we could define a random variable which takes the value 1 if the event occurs, and 0 otherwise (we'll use that "trick" sometimes in the future).*