Notes for Recitation 11 (4/24/98)

In today's tutorial we'll see how to use bijections on a probability space in order to resolve a few probability "paradoxes". We'll also see more examples on how to use conditional probabilities.

1 Predict a Diamond

Consider the following game on a deck of 10 cards with 5 diamonds and 5 clubs. The dealer shuffles the deck and starts dealing them one at a time, face up. The player tries to predict a diamond. At any point as the cards are being dealt he can say "stop!". If the next card is a diamond he wins, otherwise he loses.

What is the best strategy for the player? What is the probability that the player wins?

It is easy to find a strategy that wins half the time – predict that the first card is a diamond. From this it seems that by waiting for a few cards to be dealt and using that information we could get a better strategy. For example, we could wait till the number of clubs dealt are more than the number diamonds dealt and then predict that the next card is a diamond. Of course this may never happen for some arrangements of the deck and in such cases this strategy might lose.

Somewhat surprisingly, we will see now that in fact *no* strategy has a winning probability better than $\frac{1}{2}$. The key observation is the following:

Lemma 1.1 *After k cards have been dealt (upturned) from a shuffled deck of n cards, the probability the $(k+1)^{th}$ card is a diamond is the same as the probability that the last card is a diamond.*

Proof. Let the k upturned cards in order be $A_1, \ldots, A_k$. From the set of all permutations, consider the subset which begin with $A_1, \ldots, A_k$. Since the deck is shuffled, every permutation of the n cards is equally likely and therefore each permutation that starts out as $A_1, \ldots, A_k$ has the same probability. We will partition this set of permutations into 4 classes, depending on the values of the $(k+1)^{th}$ card and the n^{th} card. (D stands for diamond and C for club.)

1. $P_1 : A_{k+1}$ is a D and A_n is a D.

2. $P_2 : A_{k+1}$ is a D and A_n is a C.
3. $P_3 : A_{k+1}$ is a C and A_n is a D.
4. $P_4 : A_{k+1}$ is a C and A_n is a C.

Now $|P_2| = |P_3|$. To see this we set up a bijection g between the permutations in P_2 and the permutations in P_3. Let $(x_{k+1}, \ldots, x_n)$ be a permutation in P_2, so x_{k+1} is a diamond and x_n is a club. Then define

$$g((x_{k+1}, x_{k+2}, \ldots, x_n)) = (x_n, x_{n-1}, \ldots, x_{k+1})$$

So the mapping reverses the permutation. Why is this a bijection? It is an injection because two different elements of P_2 are mapped to different elements of P_3. Further, every element of P_3, say $(x_{k+1}, \ldots, x_n)$, with x_{k+1} a club and x_n a diamond, has at least one element of P_2 that is mapped to it by g, namely $(x_n, x_{n-1}, \ldots, x_{k+1})$. Hence g is also a surjection, and therefore a bijection. From this $|P_2| = |P_3|$.

Number of permutations where A_{k+1} is a D $= |P_1| + |P_2|$.

Number of permutations where A_n is a D $= |P_1| + |P_3| = |P_1| + |P_2|$.

Since all outcomes (i.e., permutations) have the same probability, we can conclude that for any outcome of $A_1, \ldots, A_k$, i.e., after dealing k cards, $Pr(A_{k+1} \text{ is } D) = Pr(A_n \text{ is } D)$. ∎

Now consider some strategy S, and set up a sample space for the outcomes of the game for that strategy. An outcome of the game is represented by a permutation of the cards and the number of cards that are dealt before the player says stop (this could be zero). Partition the outcomes according to the number of upturned cards, i.e., into n disjoint sets (corresponding to the sequence of cards that are revealed before the player says "stop"). For each of these sets separately, from the lemma, we have that the probability S wins = probability the last card is a diamond. Hence, adding up over all the outcomes for this strategy,

$$Pr(S \text{ wins}) = Pr(A_n \text{ is } D)$$

With an equal number of diamonds and clubs in the deck (5 diamonds and 5 clubs in our case), $Pr(A_{10} \text{ is } D) = \frac{1}{2}$, and hence the winning probability of any strategy (provided we say "stop" at some point) is $\frac{1}{2}$.

2 Will the Sun Rise Tomorrow?

A box contains 100 balls, some white and some black ones, possibly all white or all black ones. We assume that all the possible assortments have the same probability, namely 1/101, since there are 101 possibility (0 white balls, 1 white ball, ..., 100 white balls.)

We draw 20 balls at random from the box without replacement, and they turn out to be all white. What is the probability that if we draw a 21-st ball out of the box it will be white as well? The Bayes rule tells us that

$$Pr(21 \text{ white}|20 \text{ white}) = \frac{Pr(21 \text{ white})}{Pr(20 \text{ white})}$$

therefore the problem reduces to compute the probability that drawing 20 (respectively, 21) balls out of the bin we get all white balls.

In a more abstract formulation, we have n balls in the box; for any $i = 0, \ldots, n$ there is a probability $1/(n+1)$ that there are exactly i white balls. We draw k balls out of the box, what is the probability that they are all white? Somewhat surprisingly we will see that the answer does not depend on n.

To make the computation, we first use the rule for conditional probability.

$$\begin{aligned}
Pr(k \text{ white}) &= Pr(k \text{ white}|0 \text{ white in the box})Pr(0 \text{ white in the box}) \\
&\quad + \ldots \\
&\quad + Pr(k \text{ white}|n \text{ white in the box})Pr(n \text{ white in the box})
\end{aligned}$$

Of course the conditional probability is zero whenever we condition on the box having less than k white balls. We deduce the formula

$$Pr(k \text{ white}) = \sum_{m=k}^{n} Pr(k \text{ white}|m \text{ white in the box})Pr(m \text{ white in the box})$$

To evaluate the conditional probability, we have to define the sample space and identify the events we are interested in. The sample space is given by the $\binom{n}{k}$ ways of drawing k balls out of the box. The events whose probability we want to measure are the $\binom{m}{k}$ choices of k balls out of the m white ones. Therefore the expression becomes

$$Pr(k \text{ white}) = \sum_{m=k}^{n} \frac{\binom{m}{k}}{\binom{n}{k}} \cdot \frac{1}{n+1} \tag{1}$$

The summation of the binomial coefficients can be evaluated with the following result.

Lemma 2.1 *For any k and n such that $n \geq k \geq 0$,*

$$\sum_{m=k}^{n} \binom{m}{k} = \binom{n+1}{k+1}$$

Proof. We prove the lemma by induction on n.

Base Case: $n = 0$. Then also $k = 0$ and the statement of the lemma simplifies to

$$\binom{0}{0} = \binom{1}{1}$$

which is true.

Inductive Step: We distinguish two cases. If $n = k$ then the statement of the lemma simplifies to

$$\binom{n}{n} = \binom{n+1}{n+1}$$

which is true. Otherwise, we proceed as follows:

$$\begin{array}{rcll} \binom{n+1}{k+1} & = & \binom{n}{k} + \binom{n}{k+1} & \text{By Pascal's equality} \\ & = & \binom{n}{k} + \sum_{m=k}^{n-1} \binom{m}{k} & \text{By Inductive Hypothesis} \\ & = & \sum_{m=k}^{n} \binom{m}{k} & \text{QED} \end{array}$$

■

We can now simplify Equation 1

$$Pr(k \text{ white}) = \sum_{m=k}^{n} \frac{\binom{m}{k}}{\binom{n}{k}} \cdot \frac{1}{n+1} = \frac{\binom{n+1}{k+1}}{\binom{n}{k}} \frac{1}{n+1} = \frac{1}{k+1}$$

We can now go back to our problem and compute

$$Pr(\text{first 21 all white} \mid \text{first 20 all white}) = \frac{Pr(\text{first 21 all white})}{Pr(\text{ first 20 all white})} = \frac{1/22}{1/21} = \frac{21}{22}$$

Observe that the fact that there were 100 balls to begin with is totally irrelevant. We would have got the same result assuming there were $n \geq 21$ balls in the box and that all assortment of colors were equally likely.

Questionable extrapolations can be made out of this result. This is a classical example: if we see the same person 20 days in a row on the subway on our way to lectures then there is a 21/22 probability that we will meet him tomorrow. However if the guy gets down at Kendall, and carries an MIT folder then we would say that chances are higher. If it is the last day of lectures, then chances are lower. In general a real-world situation hardly fits the model where all possibilities are a priori equally likely.

The following even more arguable argument has been made sometimes: the sun has certainly been rising every morning in the last 27 centuries or so, since otherwise such a remarkable event as the sun not rising would have been recorded in several places and we would know about that. This sums up to about a million days, therefore the chance that the sun is not going to rise tomorrow is at most 10^{-6}. This makes no sense since we know that things do not work out so that at some initial stage it was decided how many days were going to have the sun rising and which ones were not, with all the chances being equally likely.

3 Who is Right and Who is Wrong?

10% of the problems on quizzes contain errors. If you ask a TA whether a problem has an error, he or she will answer correctly 80% of the time. If you ask the lecturer, he will answer correctly 75% of the time. The lecturer answers correctly or not independently of whether the TA answers correctly or not.

Consider now the following scenario: you have a doubt about a problem and you go to the TA asking about that, and he or she tells you that the problem has an error. To double check, you go the lecturer, and he says that the problem is correct. What is the probability that there is an error? In particular, is our belief that there is an error increased or decreased by the contradictory answers of the TA and the lecturer?

Let E be the event "the problem has an error", T be the event "the TA says the problem has an error", and L be the event "the lecturer says the problem has an error".

We want to compute $Pr(E \mid T, \overline{L})$. We use Bayes rule

$$Pr(E \mid T, \overline{L}) = Pr(E)\frac{Pr(T \cap \overline{L} \mid E)}{Pr(T \cap \overline{L})}. \tag{2}$$

Since the correctness of the answers of the TA and the lecturer are independent, we have $Pr(T \cap \overline{L}|E) = Pr(T|E)Pr(\overline{L}|E)$ and $Pr(T \cap \overline{L}|\overline{E}) = Pr(T|\overline{E})Pr(\overline{L}|\overline{E})$.

Let's compute these probability:

$$Pr(T \cap \overline{L}|E) = Pr(T|E)Pr(\overline{L}|E) = (8/10)(25/100) = 2/10$$

and

$$\begin{aligned} Pr(T \cap \overline{L}) &= Pr(T \cap \overline{L}|E)Pr(E) + Pr(T \cap \overline{L}|\overline{E})Pr(\overline{E}) \\ &= (80/100)(25/100)(1/10) + (20/100)(75/100)(9/10) \\ &= \frac{155}{1000} \end{aligned}$$

Substituting these values in equation 2 we get

$$Pr(E \mid T, \overline{L}) = \frac{1}{10} \cdot \frac{\frac{2}{10}}{\frac{155}{1000}} = .12903$$

So, even if the lecturer denies that there is an error, your confidence that there is an error increased from 10% to almost 13%.

Suppose now that when you go the TA, he or she says that the problem contains no error. Then you go to the lecturer and he admits that indeed there is an error. What is the probability of error now? Is it lower or higher than before? Conceivably it will be lower since you trust the TA better than the lecturer, but will it be *even lower* than it was before you starting asking around?

This time wee want to compute $Pr(E \mid \overline{T}, L)$. By Bayes rule

$$Pr(E \mid \overline{T}, L) = Pr(E)\frac{Pr(\overline{T} \cap L \mid E)}{Pr(\overline{T} \cap L)}. \tag{3}$$

Since the correctness of the answers of the TA and the lecturer are independent, we have $Pr(\overline{T} \cap L|E) = Pr(\overline{T}|E)Pr(L|E)$ and $Pr(\overline{T} \cap L|\overline{E}) = Pr(\overline{T}|E)Pr(L|\overline{E})$.

We now substitute the numerical values.

$$Pr(\overline{T} \cap L|E) = Pr(\overline{T}|E)Pr(L|E) = (2/10)(75/100) = 15/100$$

and

$$\begin{aligned} Pr(\overline{T} \cap L) &= Pr(\overline{T} \cap L|E)Pr(E) + Pr(\overline{T} \cap L|\overline{E})Pr(\overline{E}) \\ &= (2/10)(75/100)(1/10) + (8/10)(25/100)(9/10) \\ &= \frac{195}{1000} \end{aligned}$$

Substituting these values in equation 3 we get

$$Pr(E \mid T, \overline{L}) = \frac{1}{10}\frac{\frac{15}{100}}{\frac{195}{1000}} = .076923\ldots$$

So your confidence in the TA's opinion is such that after the TA says that the problem is OK, even if the lecturer admits an error, your belief that the problem has an error goes down to about 7.7%, which is less than the original 10%.

The calculations confirm the common sense rule that when two people make contradictory statements our belief is more influenced by what the most "authoritative" person (the one that is more often right) says. This does not mean that you necessarily *believe* in what the most authoritative person says. In the previous example, the TA's opinion makes your estimation of the probability of an error go from 10% to 13%, but still the probability is much below 50%!

4 Team problems

Problem 1. You toss a coin twice. What is the probability that you get two heads given that you get *at least one head*. What is the probability that you get two heads given that *the first toss* gives a head?

Problem 2. You generate a sequence of eight numbers $n_1, \ldots, n_8$, with $1 \leq n_i \leq 6$ for $i = 1, \ldots, 8$ in the following way. You first choose at random a position $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$ and you set $n_i = 1$. Then for any $j \neq i$ you roll a die and you set n_j according to the outcome of the die. Observe that any sequence of 8 numbers that contains at least one 1 can be generated in this way. Is the distribution described above uniform, namely are all the sequences with at least one 1 equally likely to be generated?

(a) Compute the probability of any given sequence by describing the sample space and showing which events lead to the generation of the given sequence.

(b) Which sequence(s) is/are most likely to be generated?

Problem 3. Consider the scenario of the previous section: a problem in a quiz contains an error with probability 10%, the TA has a probability 80% of answering correctly a question of yours about whether there is an error or not; the lecturer answers correctly your question with probability 75%; the correctness of the answer of the lecturer is independent of the correctness of the answer of the TA.

Is the event "the TA says that there is an error" independent of the event "the lecturer says that there is an error"?

Solutions

Problem 1

Let the sample space be $\{HH, HT, TH, TT\}$, with the four events being equally likely. So the probability of having *at least one head* is 3/4 while the probability that *the first toss is a head* is 1/2.

$$
\begin{aligned}
Pr(HH | \text{ at least one head }) &= \frac{Pr(HH)}{Pr(\text{ at least one head })} \\
&= \frac{1/4}{3/4} \\
&= \frac{1}{3}
\end{aligned}
$$

and

$$
\begin{aligned}
Pr(HH| \text{ first toss is head }) &= \frac{Pr(HH)}{Pr(\text{ first toss is head })} \\
&= \frac{1/4}{1/2} \\
&= \frac{1}{2}
\end{aligned}
$$

Problem 2

The sample space can be described by the set of events $(i; a_1, a_2, \ldots, a_7) \in \{1, \ldots, 8\} \times \{1, \ldots, 6\}^7$ where i is the choice of the place where to put a 1 and $a_1, \ldots, a_7$ are the outcomes of the rolls of the die. For example the event $(3; 2, 6, 5, 2, 1, 4, 3)$ describes the sequence $(2, 6, 1, 5, 2, 1, 4, 3)$ since we put a 1 in the third position and then we fill the other positions according to the outcomes of the die. The same sequence can be generated by more than one event. For example $(4, 1, 3, 1, 5, 6, 2)$ is generated by $(2; 4, 3, 1, 5, 6, 2)$ and by $(4; 4, 1, 3, 5, 6, 2)$. In general, let $(n_1, \ldots, n_8)$ be a given sequence, let k be the number of 1's in the sequence. We want to compute the probability of the sequence. The relevant events are all the ones of kind $(i; a_1, \ldots, a_7)$ where $n_i = 1$ and $(a_1, \ldots, a_{i-1}, 1, a_{i+1}, \ldots, a_7) = (n_1, \ldots, n_8)$. Observe that once we fix i such that $n_i = 1$ (and we have k choices for this) the values $a_1, \ldots, a_7$ are completely determined. Therefore $(n_1, \ldots, n_8)$ can be generated by precisely k different events of the sample space.

This shows that the distribution is not uniform: a sequence with more 1s is more likely to be generated and $(1, 1, 1, 1, 1, 1, 1, 1)$ is the sequence having the highest probability of being generated.

If a sequence has k 1s, then its probability is equal to k divided by the number of events in the sample space. There are 8×6^7 events in the sample space, therefore the probability is $k/(8 \times 6^7)$.

To double-check that this is a distribution, consider that there are $\binom{8}{k} 5^{8-k}$ sequences with exactly k 1s. Then we have to verify that

$$
\sum_{k=1}^{8} \frac{k}{8 \times 6^7} \binom{8}{k} 5^{8-k} = 1 \tag{4}
$$

We first observe that

$$
(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}
$$

and so

$$\frac{d((x+y)^n)}{dx} = \sum_{k=1}^{n} k \binom{n}{k} x^{k-1} y^{n-k}$$

and since

$$\frac{d((x+y)^n)}{dx} = n(x+y)^{n-1}$$

we get

$$\sum_{k=1}^{n} \frac{k}{n(x+y)^{n-1}} \binom{n}{k} x^{k-1} y^{n-k} = 1$$

which gives (4) if we set $n = 8$, $x = 1$ and $y = 5$.

Problem 3

The answer is no. We will show that the probability that the two events occur is different from the product of the probabilities. Define the events T, L, and E as before.

The probability of T is

$$\begin{aligned} Pr(T) &= Pr(T \cap E) + Pr(T \cap \overline{E}) \\ &= \frac{80}{100}\frac{10}{100} + \frac{20}{100}\frac{90}{100} \\ &= \frac{26}{100} \\ &= .26 \end{aligned}$$

The probability of L is

$$\begin{aligned} Pr(L) &= Pr(L \cap E) + Pr(L \cap \overline{E}) \\ &= \frac{75}{100}\frac{10}{100} + \frac{25}{100}\frac{90}{100} \\ &= \frac{30}{100} \\ &= .3 \end{aligned}$$

The probability of $L \cap T$ is

$$
\begin{aligned}
Pr(L \cap T) &= Pr(L \cap T \cap E) + Pr(L \cap T \cap \overline{E}) \\
&= Pr(L \cap T|E)Pr(E) + Pr(L \cap T|\overline{E})Pr(\overline{E}) \\
&= Pr(L|E)Pr(T|E)Pr(E) + Pr(L|\overline{E})Pr(\overline{E})Pr(pT|\overline{E})Pr(\overline{E}) \\
&= \frac{75}{100}\frac{80}{100}\frac{10}{100} + \frac{25}{100}\frac{20}{100}\frac{90}{100} \\
&= \frac{105}{1000} \\
&= .105
\end{aligned}
$$

Which is higher than

$$Pr(L)Pr(T) = .26 \times .3 = .078$$

The intuition is that when the TA says that the problem has an error, that it is more likely to have an error, and therefore the lecturer is more likely to find the error himself.