Tutorial 11.

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 how to make fair coins from unfair coins.

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 *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 How to make a fair coin.

Suppose you have a coin of some unknown (but fixed) bias. That is, the coin comes up heads with probability p and tails with probability $1-p$. Separate tosses are mutually independent. Can we somehow get a fair coin, i.e., one with $Pr(H) = Pr(T) = \frac{1}{2}$, from this biased coin?

The following trick for doing so is attributed to Von Neumann. Toss the unknown coin twice. Consider the possible outcomes:

1. HT

   Call that a fair head (H_f).

2. TH

   Call it a fair tail (T_f).

3. HH or TT

   Repeat, i.e. toss the coin twice again.

We will now show that the above procedure simulates a fair coin from any unknown coin.

Straightforward but painful proof.

Consider the outcomes that result in H_f. They are: HT,HHHT,TTHT,...

In general we say "heads" if we have any combination of HH and TT followed by HT. The probability of a particular outcome with α HH pairs and β TT pairs (in some fixed order) followed by HT is:

$$p^{2\alpha+1}(1-p)^{2\beta+1}$$

The number of such sequences is $\binom{\alpha+\beta}{\alpha}$. Hence, the probability of H_f is

$$\sum_{\alpha\geq 0}\sum_{\beta\geq 0}\binom{\alpha+\beta}{\alpha}p^{2\alpha+1}(1-p)^{2\beta+1}$$

To evaluate this, we change variables by substituting $t=\alpha+\beta$ to get

$$Pr(H_f)=\sum_{t\geq 0}\sum_{0\leq\alpha\leq t}\binom{t}{\alpha}p^{2\alpha+1}(1-p)^{2(t-\alpha)+1}$$

$$=p(1-p)\sum_{t\geq 0}\sum_{0\leq\alpha\leq t}\binom{t}{\alpha}(p^2)^{\alpha}((1-p)^2)^{t-\alpha}$$

$$=p(1-p)\sum_{t\geq 0}(p^2+(1-p)^2)^t$$

$$=p(1-p)\frac{1}{1-(p^2+(1-p)^2)}$$

$$=\frac{p(1-p)}{2p-2p^2}$$

$$=\frac{1}{2}$$

Similarly, $Pr(T_f)=\frac{1}{2}$. Note that there is a possibility that we never see HT or TH because we get pairs HH and TT forever but the probability of such an outcome will be 0.

Clever simple proof.

There is actually an easier way to prove this result, which will save us the effort of handling messy sums for many related problems. The idea is to make a recursive formula for $Pr(H_f)$. This is done as follows.

Consider the 4 cases for the first two coin tosses: HH,HT,TH,TT. By the Theorem of total probability (from the last tutorial),

$$Pr(H_f)=Pr(H_f\cap HT)+Pr(H_f\cap TH)+Pr(H_f\cap HH)+Pr(H_f\cap TT)$$

The following lemma will be useful in evaluating the terms in the above formula:

Lemma 2

$$Pr(H_f \mid HH) = Pr(H_f \mid TT) = Pr(H_f)$$

Proof. If we start with HH or TT, then we simply repeat the experiment. The outcome of the original experiment, is the same as the outcome of the repeated experiment. In other words, $w \in H_f$ iff $HHw \in H_f$ iff $TTw \in H_f$ for any string w. Hence,

$$Pr(H_f \mid HH) = Pr(H_f \mid TT) = Pr(H_f)$$

■

We can now compute each term separately:

$$Pr(H_f \cap HT) = Pr(H_f \mid HT)Pr(HT) = 1 \cdot p(1-p) = p(1-p)$$

$$Pr(H_f \cap TH) = Pr(H_f \mid TH)Pr(TH) = 0 \cdot (1-p)p = 0$$

$$Pr(H_f \cap HH) = Pr(H_f \mid HH)Pr(HH) = Pr(H_f) \cdot p^2 = Pr(H_f)p^2$$

$$Pr(H_f \cap TT) = Pr(H_f \mid TT)Pr(TT) = Pr(H_f) \cdot (1-p)^2 = Pr(H_f)(1-p)^2$$

Substituting these values in the formula for $P(H_f)$, we have

$$P(H_f) = p(1-p) + Pr(H_f)(p^2 + (1-p)^2)$$

$$\rightarrow Pr(H_f)(1 - (2p^2 + 1 - 2p)) = p(1-p)$$

$$\rightarrow Pr(H_f) = \frac{p(1-p)}{2p - 2p^2} = \frac{1}{2}$$

Similarly $Pr(T_f) = \frac{1}{2}$.

So we can "manufacture" a fair coin from an unknown coin. Suppose what we need is a coin of bias $\frac{1}{4}$, i.e., $Pr(H) = \frac{1}{4}$, $Pr(T) = \frac{3}{4}$. Can we simulate this using a fair coin? Yes, we could toss the fair coin twice; call it a $\frac{1}{4}$-head if it comes up head both times, and a $\frac{3}{4}$-tail for all other outcomes. It is easy to see that this has the required probability.

What about any q-coin? That is, can we simulate a coin which comes up head with probability q and tails with probability $1-q$ for any given value of q between 0 and 1?

Let the binary expansion of q be $0.b_1b_2b_3\ldots$ where b_i is the i^{th} bit and it is 0 or 1. Let 0 correspond to a a tail (T) and 1 to a head (H). Now consider the following experiment:

1. Start tossing the fair coin, once for each bit after the decimal point in the binary expansion of q.

2. Stop when the outcome of the coin toss agrees with the corresponding bit of the expansion, i.e., stop at the smallest i for which the i^{th} coin comes up heads and $b_i = 1$ or it comes up tails and $b_i = 0$.

3. If it stops on an H, i.e., the coin comes up H on a bit that is 1, call it a head for the q-coin, otherwise call it a tail.

Lemma 3 *The probability of a head in the above experiment is q.*

Proof. The experiment stops when the outcome of the fair coin agrees with the corresponding bit of the binary expansion of q. The probability the experiment stops on the first toss is $\frac{1}{2}$. The probability the experiment stops after i tosses = probability the coin disagrees with the first $i-1$ bits and agrees with the i^{th} bit $= (\frac{1}{2})^i$. Whenever the experiment stops at a bit that is 1 we call it a head for the fair coin. Therefore, to calculate the total probability that we stop and say "head" we have to add up the probabilities of stopping at any bit that is 1 in the expansion of q.

Formally, the probability of stopping and saying head =

$$\begin{aligned} & Pr(\cup_{i:b_i=1} \text{stopping at } b_i) \\ = & \sum_{i:b_i=1} Pr(\text{stopping at } b_i) \end{aligned}$$

We can use a simple sum (instead of inclusion-exclusion) since the the events of stopping at two different b_i are disjoint (intersection is empty). Thus,

$$\begin{aligned} Pr(q\text{-head}) &= \sum_{i=1}^{\infty} b_i(\frac{1}{2})^i \\ &= q \end{aligned}$$

■

Note that since we can simulate a fair coin from an unknown coin and a coin of any bias from a fair coin, we can also simulate a coin of any bias from a coin of unknown bias.

3 How to make a fair coin — Not!

Consider the following attempt to make a fair coin from a coin of unknown bias p: We toss the coin till we get HT or TH and then call it a head (H_f) if HT occurs before TH and a tail if (T_f) if TH occurs before HT.

This seems more efficient than breaking up a sequence of tosses into pairs since we will get a result sooner, but it does *not* give $Pr(H_f) = \frac{1}{2}$. Actually it just gives $Pr(H_f) = p$, which is no better than before. Here is why:

$$\begin{aligned} Pr(H_f) &= Pr(H_f \cap H) + Pr(H_f \cap T) \\ &= Pr(H_f \mid H)Pr(H) + Pr(H_f \mid T)Pr(T) \end{aligned}$$

Now

$$Pr(H_f \mid H) = 1 - Pr(\bar{H_f} \mid H)$$

But $Pr(\bar{H_f} \mid H) =$ probability of an infinite string of H's $= \lim_{n \to \infty} \frac{1}{2^n} = 0$ Hence

$$Pr(H_f \mid H) = 1$$

Similarly, $Pr(H_f \mid T) = 0$. Then from the above formula we have

$$Pr(H_f) = Pr(H) = p$$

Similarly, $Pr(T_f) = Pr(T) = 1 - p$.

4 Team problems

1. If you repeatedly flip a fair coin, what is the probability you see the sequence HTT before you see the sequence HHT (where H is a head and T is a tail)? [Hint: First evaluate the conditional probability of seeing HHT before HTT given that that the first toss comes up H. Then show that this is the same as the required probability. The answer is not $\frac{1}{2}$!]

2. If you repeatedly flip a fair coin as above, what is the probability that you see the sequence HHT before you see the sequence TTH ? [Hint: Your answer and the proof should be very simple.]