EE 178/278A Probabilistic Systems Analysis
Spring 2014 Tse/Hussami Lecture 5

Biased Coin Review

Suppose we have two biased coins, with probabilities p for coin 1 and q for coin 2 of obtaining heads.

We first consider the experiment where we generate a sequence of length 5 by each time randomly picking one of the two coins and then flipping it. Let B_i be the event that the ith flip is a heads. Then the events B_i are independent. We can compute:

$$\mathbf{P}(HHTTT) = \mathbf{P}(H)^2\mathbf{P}(T)^3$$

where $\mathbf{P}(H) = \frac{1}{2}p + \frac{1}{2}q$ by the total probability rule. (Note also that the probability of this outcome depends only on the number of heads and tails it has and not on the ordering.)

However, if we select a coin initially, and then flip that same coin 3 times, the events B_i are not independent anymore. The identity of the coin is hidden from us in this case. However, if we condition on which coin was chosen, then the flips are independent (it is as if we were in an experiment where we only have the chosen coin).

Let A be the event that coin 1 is chosen in the beginning. Then

$$\mathbf{P}(HHTTT) = \mathbf{P}(HHTTT|A)\mathbf{P}(A) + \mathbf{P}(HHTTT|A^c)\mathbf{P}(A^c)$$

where $\mathbf{P}(A) = \mathbf{P}(A^c) = \frac{1}{2}$ and $\mathbf{P}(HHTTT|A) = \mathbf{P}(H|A)^2\mathbf{P}(T|A)^3$.

Conditional Independence: We define events A and B to be conditionally independent given C if

$$\mathbf{P}(A \cap B|C) = \mathbf{P}(A|C)\mathbf{P}(B|C)$$

The notion of conditional independence is very important. In many problems, events are not independent but they become independent when conditional on a third event. This third event is often about something we want to infer about and therefore "hidden". Consider again the medical diagnosis example. Suppose we are not satisfied with the accuracy of the test. Then perhaps we should use an additional test, and combine the results from the two tests to come up with a diagnosis. Now the results of the two tests are not independent, since they both depend on whether the patient has the disorder or not. On the other hand, if the tests complement each other well, then their results should be independent given the health of the patient. (Is the additional test useful if its result is always identical to the first test?)

Example: The coupon collection problem[1]

Suppose that when we buy a box of cereal, as a marketing ploy, the cereal manufacturer has included a random baseball card. Suppose that there are n baseball players who appear on some card, and each cereal

[1] This section is adapted from Berkeley EECS 70 notes.

box contains a card chosen uniformly and independently at random from these n possibilities. Assume that Babe Ruth is one of the n baseball players, and I am a huge fan of him: I really want his baseball card. How many boxes of cereal will I have to buy, to have at least a 90% chance of obtaining a Babe Ruth card? This problem can be analyzed as follows. Suppose we buy m boxes of cereal. Let E be the event that I do receive a Babe Ruth card in any of the m boxes. The card in each of the m boxes is independent, and for each box the chances that I don't receive Babe Ruth's card from that box is $1-\frac{1}{n}$. Suppose, that we want to collect all n cards with 90% chance. How many boxes do we need to buy in order to achieve that?

We first find m such that $\mathbf{P}(E) \geq 0.9$. E^c is the event that none of the m boxes contain BR. Let F_i be the event that the *ith* box does not have Babe Ruth. We can write $E^c = F_1 \cap F_2 \dots F_m$.

Now, we have that

$$\mathbf{P}(E^c) = \mathbf{P}(F_1 \cap F_2 \cdots \cap F_m) = \mathbf{P}(F_1)\dots\mathbf{P}(F_m) = \mathbf{P}(F_1)^m$$

where $\mathbf{P}(F_i) = 1-\frac{1}{n}$. Therefore, $\mathbf{P}(E) = 1-(1-\frac{1}{n})^m$.

We would like to get $\mathbf{P}(E) \geq 0.9$, so we get

$$1-(1-\frac{1}{n})^m = 0.9 \qquad \Longrightarrow (1-\frac{1}{n})^m = 0.1 \qquad \Longrightarrow m = \frac{\log(0.1)}{\log(1-\frac{1}{n})} \tag{1}$$

hence $m \approx n\log(10)$. Note that m depends linearly on n: if there are twice as many cards, you need to buy twice as many boxes in order to guarantee with 90% chance you will get Babe Ruth.

Next suppose that what I really want is a complete collection: I want at least one of each of the n cards. Let A be the event that we get the whole collection after m boxes. Then, A^c is the event that we are missing at least one of the cards after m boxes. This event can be broken down into simpler events. How is A^c related to E?

Lets give a number to each player and suppose Babe Ruth is player number 1 (who else?). Let E_1^c be the event that BR is not among the cards in the m boxes. Similarly, we let E_i be the event that player i is among the cards in the m boxes and E_i^c is the event that player i is not among the cards in the m boxes. A^c can be expressed in terms of E_i as

$$A^c = E_1^c \cup \dots E_n^c$$

Now, by symmetry,

$$\mathbf{P}(E_i^c) = (1-\frac{1}{n})^m, \qquad i \in 1,\dots,m.$$

We are interested in finding an m that is large enough such that $\mathbf{P}(A^c) \leq 0.1$. Therefore, we only need to bound $\mathbf{P}(A^c)$ from above.

Using the union bound (i.e. $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$ for any events A and B), we get that

$$\mathbf{P}(A^c) \leq \mathbf{P}(E_1^c) + \cdots + \mathbf{P}(E_n^c) = n(1-\frac{1}{n})^m.$$

Therefore it is sufficient to solve for a value of m such that $n(1-\frac{1}{n})^m = 0.1$, which gives $m \approx n\log(10n)$.

In the case of collecting a specifc card (say Babe Ruth), the number of boxes one needs to collect grows linearly with the number of cards. In the case of collecting a whole collection, however, the number of boxes needed grows *super-linearly* with the number of cards. This means that if there are twice as many cards to collect, you need to buy twice the number of boxes to collect them. This is because it becomes more and more difficult to get a new card once we have already collected several cards.

Perhaps the most interesting aspect of the coupon collector's problem above is that it illustrates the use of the union bound to upper-bound the probability that something bad happens. In this case, the bad event is that we fail to obtain a complete collection of baseball cards, but in general, this methodology is a powerful way to prove that some bad event is not too likely to happen.