OMCIT 592 Module 07 Self-Paced 02 (instructor Val Tannen)

Reference to this self-paced segment in seg.07.04 slide 6 (actually…... not, but it could have been!)

This is a segment that contains material meant to be learned *at your own pace*. We are trying to assist you in this endeavor by organizing the material in a manner similar to the way it is outlined in the recorded segments, however with one additional suggestion.

When you see the following marker:

we suggest that you stop and make sure you thoroughly understood the material presented so far before you proceed further.

Marbles from an urn

In the lecture segment "Birthdays, balls, and bins", we explained the Birthday Paradox and formulated a more general probabilistic model that we called Balls into Bins. To help you approach a larger variety of different problems based on probabilistic models, we formulater here another such model which we will call Marbles from an Urn.

In this probabilistic model we have an opaque urn that contains marbles of equal size and weight but of different colors. Marbles are extracted (sampled) from the urn and each marble is equally likely to be taken, as the color is only discovered after extraction.

To keep the notation simple let's assume that we have just tree colors, specifically that the urn contains r red marbles, b blue marbles and g green marbles. Let also $n = r + b + g$.

We have **two alternatives** for modeling a **single extraction**.

The **first alternative** is to use a **uniform** probability space with n outcomes, one for each marble, and to model the colors as specific events in this space: R (Red), B (blue), and G (green) with corresponding probabilities $\Pr[R] = r/n$, $\Pr[B] = b/n$, and $\Pr[G] = g/n$.

Note that the events R, B, G are pairwise disjoint and their union is the entire probability space. Therefore, by probability properties (P2gen) and (P1) we have:

$$\Pr[R] + \Pr[B] + \Pr[G] = 1$$

Sanity check:

$$r/n + b/n + g/n = (r + b + g)/n = n/n = 1$$

Then, events like "the marble extracted is not green" and "the marble extracted is red or blue" can be expressed as follows:

- By (P4):

$$\Pr[\text{"the marble extracted is not green"}] = 1 - \Pr[G] = 1 - g/n.$$

- By (P2):

$$\Pr[\text{"the marble extracted is red or blue"}] = \Pr[R] + \Pr[B] = r/n + b/n = \frac{r + b}{n}$$

As a sanity check we show that the two probabilities are actually the same: $\frac{r+b}{n} = \frac{n-g}{n} = 1 - g/n$.

(CONTINUED)

Marbles from an urn (continued)

The **second alternative** for modeling a **single extraction** is to associate outcomes with colors. This leads to a non-uniform probability space with **three outcomes**:

1. "The marble extracted is red" with probability r/n,

2. "The marble extracted is blue" with probability b/n,

3. "The marble extracted is green" with probability g/n.

This is a valid probability space because $r/n + b/n + g/m = (r + b + g)/n = n/n = 1$.

It is easy to see that events like "not green" and "red or blue" have the same probability as in the first alternative. (Try it!)

The first alternative (artificially) distinguishes the marbles as outcomes but its advantage is that the space it produces is uniform.

The second alternative reflects the fact that marbles are only distinguishable by their color and seems more intuitive.

We should point out that the second alternative is a completely general model, in that *any* finite probability space can be seen in this way. One can then move to the first alternative and use a uniform space. In fact, this is exactly what we did in the lecture segment "Biased coins and Bernoulli trials" in order to use a uniform space to compute a probability related to a non-uniform biased coin.

(CONTINUED)

Marbles from an urn (continued)

Until now we have been discussing **single marble extractions**. What about if we extract 2 marbles **simultaneously**? Let's consider the following problem:

Problem. An urn contains r red marbles, b blue marbles, and g green marbles such that $n = r + b + g \geq 2$. We extract 2 marbles **simultaneously**. Calculate the probability that the marbles have different colors.

Answer. Since the two alternatives discussed above correspond to the extraction of a single marble, we need another probability space here that will formalize what it means to extract two marbles simultaneously.

In the spirit of the first alternative discussed above we assume that any subset of of 2 marbles is equally likely to be extracted, and take these as the outcomes. This gives a uniform probability space with $\binom{n}{2}$ outcomes.

The event E of interest consists of the subsets of 2 marbles of different colors. Now we only need to count $|E|$. We do this by first counting the number of sequences of length 2 consisting of marbles of different colors. By the addition and multiplication rules this is

$$rb + rg + br + bg + gr + gb = 2(rb + bg + gr)$$

Next we must divide by 2 because each subset of interest can be ordered in 2 ways to give a sequence. Hence $|E| = rb + bg + gr$.

The probability we seek is:

$$\frac{|E|}{|\Omega|} = \frac{rb + bg + gr}{\binom{n}{2}} = \frac{2(rb + bg + gr)}{n(n-1)}$$