CMI/BVR Probability notes 2

Ballot Problem:

Recall K got 60 votes and S got 40 votes. Assuming all ‘vote countings’ are equally likely we want to find out the chances that K is leading through out.

A counting is a path starting at $(0,0)$ and ending at $(100,20)$. Number of paths starting at $(0,0)$ and ending at (n,r) are

$$\binom{n}{\frac{n+r}{2}} = \binom{n}{\frac{n-r}{2}}.$$

Thus total number of countings are $\binom{100}{60}$. In the formula above if $(n+r)/2$ is not an integer (≥ 0) then its value is by definition, zero.

Let us note a simple but powerful fact.

Reflection Principle:

■ Let $k \geq 1$ and $r \geq 1$ be integers.
number of paths from $(0,k)$ to (n,r) which touch or cross the x-axis
equals
number of paths from $(0,-k)$ to (n,r).

Note $(0,-k)$ is reflection of $(0,k)$ in the x-axis. The proof of this fact uses reflection in x-axis.

Let A be the first set of paths and B be the second set of paths. Take a path π in A.

$$(0, s_0), (1, s_1), \cdots\cdots (n, s_n); \qquad s_0 = k; \quad s_n = r.$$

Since it touches the x-axis; let i be the first index with $s_i = 0$. Reflect the path till that point in x-axis. do not reflect the remaining segment of the path. That is, consider the path π^*:

$$(0, -s_0), (1, -s_1), \cdots (i, s_i), (i+1, s_{i+1}), (i+2, s_{i+2}), \cdots (n, s_n).$$

Remembering that $s_i = 0$, we see (i, s_i) is same as $(i, -s_i)$. We can easily verify that this is also a path, that is, successive s-differences are ± 1. This path starts at $(0, -k)$ and ends at (n, r). Hence $\pi^* \in B$.

Given path $\eta \in B$ it starts below x-axis and ends above; so must hit x-axis, take the first time it hits and reflect the part till then in x-axis, keeping later part as is. This will be a path $\pi \in A$ and $\eta = \pi^*$. and actually the map $\eta \mapsto \pi$ is inverse map to $\pi \mapsto \pi^*$.

This one-one map between A and B proves the result. ∎

We are interested in
number of paths $(0,0)$ to $(100, 20)$ that do not touch/cross x-axis.
[such a path should pass through $(1,1)$]

= number of paths from $(1,1)$ to $(100, 20)$ that do not touch x-axis
[subtracting 1 from first coordinate to see]

= paths from $(0,1)$ to $(99, 20)$ that do not etc

= [paths $(0,1)$ to $(99,20)$] minus [paths $(0,1)$ to $(99,20)$ that touch etc]
(Use reflection principle)

=[paths $(0,1)$ to $(99,20)$] minus [paths $(0,-1)$ to $(99,20)$]

= [paths $(0,0)$ to $(99,19)$] minus [paths $(0,0)$ to $(99,21)$]

$$= \binom{99}{40} \text{minus} \binom{99}{39}$$

$$= \binom{100}{40} \frac{60-40}{100}.$$

Since the total number of outcomes is $\binom{100}{40}$ we conclude that the required probability is $20/100$ as stated.

Birth day problem:

Consider $n = 29$ students of this class. assume that birth day of each student could be any of the the 365 days, so that there are $(365)^{29}$ possible outcomes as birthday sequence of these 29 students. We assume that these are equally likely.

What are the chances that no two birthdays are equal. In other words what are the chances that all birth days are different? clearly it equals

$$p_{29} = \frac{(365)_{29}}{(365)^{29}} = 1\left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{28}{365}\right)$$

The interesting point is that already for $n = 29$, there is a very high chance that two students will have a common birthday! Thus, when the problem was first considered, people were surprised. And hence this problem is also referred to as birth-day paradox. It is more like surprise than paradox. The table of p_n values (saw some where on internet, but better if you compute) are as follows.

number of people	p_n (no two equal)	$1 - p_n$ (at least one equal pair)
1	1	0
2	0.997	0.003
3	0.992	0.008
4	0.984	0.016
5	0.973	0.027
10	0.883	0.117
15	0.747	0.253
20	0.589	0.411
25	0.431	0.569
30	0.294	0.706
35	0.186	0.814
40	0.109	0.891
50	0.030	0.970
60	0.006	0.994

Coupon collector problem:

I have 10 sets of playing cards. I pick 100 cards from these 520 cards without replacement. What are the chances that the sample contains cards to make a full deck, that is, the sample contains at least one card of each of the 52 denominations.

There are 520 cards and $(520)_{100} = N$ (say) samples. Let A be the event that the sample avoids at least one denomination. Thus the required probability equals

$$1 - P(A)$$

Let A_i be the event that denomination i is avoided in the sample. Here $i = 1, 2, \cdots, 52$. So

$$P(A) = P(\cup A_i) = S_1 - S_2 + S_3 + \cdots \pm S_{52}$$

Clearly

$$S_1 = 52(510)_{100}/N; \quad S_2 = \binom{52}{2}(500)_{100}; \quad S_3 = \binom{52}{3}(490)_{100}; \cdots$$

Substituting this in the equation above, we get the required probability.

If there are s sets each of n cards you can write down the general formula.

This is called coupon collector problem for the following reason. A company, say Bournvita, includes one coupon – like batman/spiderman/mickey mouse etc — with each Bournvita packet. Assume there are n possible caricatures. It prints s sets of these caricatures to put with its product. They announce that if you collect a full set and produce then you will receive a gift.

Thus the interesting problem is on the average how many items do you need to buy (how long you should wait) so that you acquire a full set. In general they print a very large number of sets, so that you can consider as if the sampling is with replacement. We shall return to this problem again later. This arises in other contexts too.

sticks problem:

I have 23 sticks numbered $1, 2, \cdots, 23$. Break each into two parts one small part and one large part. Now I make a bag of 23 sticks again by pairing these 46 pieces.

What are the chances of getting back origin pairing?

What is the probability that long pieces are paired with small parts?

Here

$$|\Omega| = \frac{46!}{2^{23}\,(23)!}$$

This is because, if c is the number of ways of pairing, consider the problem of pairing and arranging the pairs in a line. This can be done using two methods:

(Method 1) make pairs (c ways) and then arrange them in a row ((23)!ways). so $c\,(23)!$ ways.

(Method 2) pick a pair out of th 46 pieces, put in place one;
pick a pair out of the rest, put in place 2 etc.
so can be done in

$$\binom{46}{2}\binom{44}{2}\cdots\binom{2}{2} = \frac{46!}{2^{23}}$$

As a result

$$c\,(23)! = \frac{46!}{2^{23}}; \qquad c = \frac{46!}{2^{23}(23)!}$$

Thus each outcome has probability

$$\frac{1}{c} = \frac{2^{23}\,(23)!}{(46)!}$$

which is the probability of getting original pairing.

If A is the set of all pairings where large parts are paired with short parts, then $|A| = (23)!$ — you need to put long pieces in a line, no matter how, and then place short pieces one with each long piece. So the required probability equals

$$\frac{2^{23}\,(23)!\,(23)!}{(46)!} = \frac{2^{23}}{\binom{46}{23}}$$

As you could guess these 23 sticks are the 23 chromosomes in human cell. They break at the centromere and then join. The first question asks the chances of getting original chromosomes back or original cell back. The second question asks for the chances of the cell dying; if two long pieces join there is not enough space in the cell to fit in.

What happens if there are N sticks?

not necessarily equally likely outcomes:

Let us now discuss experiments where the outcomes may not be equally likely. Thus we have a countable sample space and assignment of probability for each outcome. Recall the following.

A good assignment of probability for points of a sample space is a map that associates with each $\omega \in \Omega$ a non-negative number $p(\omega)$ which all add to one. Then Probability of an event A equals $\sum_{\omega \in A} p(\omega)$.

The number $p(\omega)$ is the chances of observing the outcome ω.

The sample space with the assignment of probability (Ω, p) is called **probability space**.

Example: $\Omega = \{H, T\}$

$p(H) = 0.2, \qquad p(T) = 0.8$

This experiment corresponds to tossing a coin once, the coin is biased, chances of Heads are 0.2.

Example: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$p(1) = p(2) = p(3) = 0.1 \quad p(4) = p(5) = 0.2 \quad p(6) = 0.3$

This corresponds to rolling an unbiased die once.

If A is the event $\{1, 5, 6\}$ then $P(A) = 0.6$. Similarly, you can calculate probability of any event.

Example: $\Omega = \{H, TH, TTH, TTTH, \cdots\}$

$p(T^n H) = \frac{1}{2^{n+1}} \quad n = 0, 1, 2, 3, \cdots$.

This corresponds to tossing a fair coin till Heads are obtained and then stop.

If, for example $A = \{T^n H : n = 1, 3, 5, \cdots\}$ then

$P(A) = \sum_1^\infty 2^{-2n} = \frac{1}{3}$

The first question that arises is whether the earlier rules hold good. Yes.

(1) If A, B are disjoint events, $P(A \cup B) = P(A) + P(B)$.

Because, there are no outcomes in common between A and B

$$\sum_{\omega \in A \cup B} p(\omega) = \sum_{\omega \in A} p(\omega) \quad + \quad \sum_{\omega \in B} p(\omega).$$

(2) For any two events A, B; $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Indeed the same proof as in equally likely case goes through!

$$P(A \cup B) = P(A \cup [B - A \cap B]) = P(A) + P(B - A \cap B)$$

$$P(B) = P([A \cap B] \cup [B - A \cap B]) = P(A \cap B) + P(B - A \cap B)$$

Use second equation in the first one.

(3) For any events $A_1, A_2, \cdots, A_N$

$$P(\cup A_i) = S_1 - S_2 + S_3 - \cdots \pm S_N.$$

As earlier

$$S_1 = \sum P(A_i); \quad S_2 = \sum_{i<j} P(A_i \cap A_j) \quad \cdots$$

Exactly the same proof as in the equally likely case works. Remember, we use induction, for $N = 2$ done above. The inductive step was essentially rearranging terms.

This proves the formula above. Here is another proof. Both sides are sum of certain of the numbers $p(\omega)$. Take any $\omega \in \Omega$. We show $p(\omega)$ is added the same number of times (zero or one) on both sides.

If $\omega \notin \cup A_i$ then $p(\omega)$ does not appear in calculating left side. It does not appear in calculation of any of the S_i and hence does not appear on the right side either.

If ω is in exactly one A_i, say A_5. Then $p(\omega)$ appears in the calculation of only $P(A_5)$ and in none of the other. Thus it appears exactly once in calculating S_1 and does not appear in calculating other S_i. Thus $p(\omega)$ is added exactly once on right side as well as left side.

Let ω appear in exactly n of the events. let us see how many times $p(\omega)$ is added on the right side. In calculating S_1 it is added n times. In calculating S_2 it is added $\binom{n}{2}$ times and so on. Thus (remembering S_2 appears with minus sign etc) it is added on right side as many times as

$$n - \binom{n}{2} + \binom{n}{3} - \cdots \pm \binom{n}{n} \;=\; 1 - (1-1)^n.$$

Thus $p(\omega)$ is added exactly once on each side.

This proves the formula. The advantage of this proof is that it can be used to prove a more general formula.

(4) Again let $A_1, \cdots, A_N$ be events. Let $1 \le m \le N$. let E be the set of all outcomes which belong to exactly m of these events. Take those outcomes ω satisfying $|\{1 \le i \le N : \omega \in A_i\}| = m$

$$P(E) = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - \cdots \pm \binom{N}{m} S_N.$$

We earlier found out chances of at least one match and hence (by taking complements) chances of 'exactly zero' matches. The above formula will help you find chances of 'exactly 10 matches'.

Here is the proof. Take any sample point ω. We show that $p(\omega)$ is aded the same number of times on both sides.

(i) Suppose $|\{1 \le i \le N : \omega \in A_i\}| < m$. Then $\omega \notin E$; So $p(\omega)$ is not added on left side. Also on the right side, in the calculation of any of the $S_i, i \ge m$ it does not appear.

(ii) Suppose $|\{1 \le i \le N : \omega \in A_i\}| = m$. Then $\omega \in E$ and $p(\omega)$ is added on the left side. On the right side, $p(\omega)$ does not appear in the calculation of any of the $S_i, i \ge m+1$. It appears in the calculation of S_m and only once it appears when we calculate the probability of the intersection of the m sets to which ω belongs.

(iii) Suppose $|\{1 \leq i \leq N : \omega \in A_i\}| = n > m$. Then $p(\omega)$ does not appear in the calculation of LHS. Following the same idea as earlier it appears on RHS as many times as the following:

$$\begin{aligned} &\binom{n}{m} - \binom{m+1}{m}\binom{n}{m+1} + \binom{m+2}{m}\binom{n}{m+1} - \cdots \\ = &\binom{n}{m} - \binom{n}{m}\binom{n-m}{1} + \binom{n}{m}\binom{n-m}{2} - \cdots \pm \binom{n}{m}\binom{n-m}{n-m} \\ = &\binom{n}{m}(1-1)^{n-m} = 0 \end{aligned}$$

as required. This completes the proof of the formula.

conditional probability:

One of the most important concepts is that of conditional probability. Most of the time we are not totally ignorant. There is a chance experiment, we do not know the outcome but we know some partial information about the experiment.

Example: Consider rolling a fair die.
$\Omega = \{1, 2, 3, 4, 5, 6\}$ and $P(A) = |A|/6$.

Some one told us that an even number appeared. How shall we define probabilities now? Let us denote $B = \{2, 4, 6\}$. It is meaningless to say that the probabilities of events are as earlier [strictly speaking, meaningful but practically useless]. If we are now asked chances of the event $A = \{2, 3, 4\}$ then we feel that $P(A)$ should be $2/3$. Because, there are now only three possibilities: $2, 4, 6$, of which the event A has two: $2, 4$. Of course, if we say $P(A) = 2/3$ there will be total confusion because in the original experiment $P(A) = 3/6$. So we say $P(A|B) = 2/3$; conditional probability of A given B. Here 'given B' means that an outcome in B occurred.

Similarly, if tomorrow someone tells that an odd number occurred, then you need to rethink and say that $1, 3, 5$, are the possible outcomes with this information $P(A) = 1/3$ etc. It is a nuisance to keep on changing sample space and *then* calculate. One smart fellow found out the quantity $P(A|B)$ is nothing but $P(A \cap B)/P(B)$.

This formula has the advantage that it expresses the conditional probability — that we felt — as ratio of probabilities. Thus you can calculate conditional probability without any sample space considerations, using only probabilities. we take this as the general definition, whether the outcomes are equally likely or not.

In an experiment (Ω, p) *let* B *be any event. We define for every event* A*, conditional probability of* A *given* B *by* $P(A|B) = P(A \cap B)/P(B)$.

Naturally when $P(B)$ is zero, we shall not define the conditional probability. Actually, it does not matter what you do in such a case, but let us not confuse now.

Before passing let us be clear that $(A|B)$ is NOT an event. A and B are events, but not $(A|B)$. In fact this symbol $(A|B)$, standing by itself has no meaning. However $P(A|B)$ has meaning; it is the conditional probability of the event A in our experiment when we knew that B happened.

A question that naturally arises is the following. Will the discovery of the smart fellow remain reasonable even in the non-equally likely case. Yes, let us just discuss one illustration. Let us consider the following biased die.

$$p(1) = p(2) = p(3) = 0.1 \quad p(4) = p(5) = 0.2 \quad p(6) = 0.3$$

Let $B = \{2, 4, 6, \}$. Thus we are informed that an even number turned up. How do we redefine the probabilities. Obviously, *now* odd faces must have probability zero. Let $c = p^*(2)$. Since face 4 was double as likely as face 2, we should still agree $p^*(4) = 2c$ and for similar reason $p^*(6) = 3c$ and since the probabilities should add to one we must have $(c + 2c + 3c) = 1$ or $c = 1/6$. Thus the conditional probabilities should be

$$p^*(1) = p^*(3) = p^*(5) = 0.1; \quad p^*(2) = 1/6,\ p^*(4) = 2/6;\ p^*(6) = 3/6$$

If you use these conditional probabilities of outcomes — that is, use this assignment of probabilities to outcomes — to calculate probabilities of events you will see $P^*(A) = P(A \cap B)/P(B)$. Thus the definition stands to intuition even in the non-equally likely case.

Just as with any mathematical formula, the definition of conditional probability, $P(A|B) = P(A \cap B)/P(B)$, can be used to calculate any one of the three quantities if we know the other two. Thus *if we know probabilities we can calculate conditional probabilities using the formula. If we know some probabilities and some conditional probabilities then we can use them to calculate probabilities.*

Or, if we are prescribed some probabilities and some conditional probabilities, then we can assign probabilities so that these requirements are met.

Of course this last sentence is to be understood in the following spirit: if the given hypotheses (probabilities and conditional probabilities) are reasonable then we can assign probabilities. We shall not spend time on making this precise, though it is not difficult in the present set up.

Polya Urn:

Consider a box with r red balls and g green balls. here is the experiment.

Pick a ball at random; see colour; put it back;
add a ball of that colour. *Repeat.*

This is called Polya urn model.

Let R_n be the event that the n-th draw is Red. Clearly $P(R_1) = r/(r+g)$.

What is $P(R_2)$? unclear because we were not told what happened in draw one. If we are told that first draw was red then the chance of red ball now is $(r+1)/(r+g+1)$ simply because there is one more red ball now. if we are told that first draw was green then the chance of red ball now is $r/(r+g+1)$. Thus we know conditional probabilities! We are in business!

$P(R_2) = P(R_2R_1) + P(R_2G_1)$ R_2 is split into disjoint events

$= P(R_1)P(R_2|R_1) + P(G_1)P(R_2|G_1)$ def. conditional probability.

$$= \frac{r}{r+g}\,\frac{r+1}{r+g+1} + \frac{g}{r+g}\,\frac{r}{r+g+1} = \frac{r}{r+g}$$

Thus Chance of red ball remains the same. In fact it remains same for every draw as we see soon. But some thing is changing: The conditional probabilities.

$$P(R_2|R_1) = \frac{r+1}{r+g+1} > \frac{r}{r+g} = P(R_1)$$

Thus if we se a red ball then chances of seeing red ball again are increased.

This is very interesting phenomenon.

Kisalaya says he came from Lucknow. Imagine that when he was leaving Lucknow to come here a friend told him: be careful, there is flu epidemic in Chennai. So K lands at Chennai station. If he sees one fellow with flu what would he think? Oh yes, my friend is right, there are many people with flu here. If next person he sees also has flu, he tends to think: Oh my god there is lot of flu here. On the other hand the first person he meets after landing has no flu, he feels relieved and thinks, oh there is not that much flu here.

Thus in his mind he revises (chances of seeing a flu person) depending on his experience. But reality is same irrespective of K's experiences. There is a certain proportion, say p, with flu and the chance of seeing a person with flu is p. [in the small time duration we are concerned, spread of the epidemic is to be ignored and do not get confused].

Math, like music, can be improvised, once the basic notes are prescribed!

For example why add one ball of that color? Add 20 balls of that color. The phenomenon remains the same.

When K sees a person with flu, his revised opinion of seeing a flu person (or the conditional probability he attributes to seeing flu person) is $(r+20)/(r+g+20)$ which is much much larger than $(r+1)/(r+g+1)$. Thus if K is very very nervous person, then a large number like 20 is more suited than the number one.

Why add ball of that color, why not ball of opposite color? Yes, can do, this also models an interesting phenomenon; shall return to this. Of course $P(R_n)$ no longer remains same for all n.

Whay add only balls of one color? add 2 balls of the color seen and one ball of the opposite color. Yes, leads to interesting mathematics.

Why only two colors? why not start with twenty colors? Yes, can be done and has certain applications too.

Let us return to Polya scheme. For $n = 1, 2, 3 \cdots$, let S_n be the following statement.

■ whatever be $r, g \geq 1$, in the Polya model $P(R_n) = 1$

We know S_1 and S_2 are true. We prove these statements by induction and by using the exact same argument as used for $n = 2$ above. Suppose the statement S_m is true. Let us prove S_{m+1}. Start with any $r, g \geq 1$.

$$P(R_{m+1}) = P(R_{m+1}R_1) + P(R_{m+1}G_1) \quad R_{m+1} \text{ is split into disjoint events}$$

$$= P(R_1)P(R_{m+1}|R_1) + P(G_1)P(R_{m+1}|G_1) \quad \text{def. conditional probability.}$$

$$= \frac{r}{r+g}\,\frac{r+1}{r+g+1} + \frac{g}{r+g}\,\frac{r}{r+g+1} = \frac{r}{r+g}$$

We used the induction hypothesis to calculate $P(R_{m+1}|R_1)$ as follows: given R_1 the urn has $r + 1$ Red and g green balls and we want probability of red at the 'now' m-th draw. Similarly induction does for $P(R_{m+1}|G_1)$. This completes the proof. ∎

You should notice the subtle phrase: 'whatever be r, g' in our statement S_n. This was essential because we used the induction hypothesis NOT for (r, g) but for $(r + 1, g)$ and $(r, g + 1)$ in the proof.

That is why it is important for you to make clear to yourself as to what exactly is the statement that is being proved by induction.

We observe one more interesting phenomenon of this model.

$$P(R_1G_2) = P(R_1)P(G_2|R_1) = \frac{r}{r+g}\,\frac{g}{r+g}$$

$$P(G_1R_2) = P(G_1)P(R_2|G_1) = \frac{g}{r+g}\,\frac{r}{r+g}$$

Since $P(R_1) = P(R_2)$ we conclude from the above two equations

$$P(G_2|R_1) = P(G_1|R_2). \qquad (\spadesuit)$$

This is very interesting. Both sides ask chances of green; both sides give the same information namely ball is red. What is the difference? left side gives information about first draw and asks about second draw. Right side gives the same info about second draw and asks the same question about first draw. Both are equal. There is 'time symmetry'!

For a moment, think of time as 'generation' and red/green are genes — we concentrate on one of the 22 (?) chromosomes and one locus and the chemical there, to simplify matters. Thus first generation is my father and second generation is me. LHS asks the chances of my having green gene given that my father has red gene. RHS asks the chances of my father having green gene given that I have red gene. Thus what the above says is that both are equal. This is not as ridiculous as it appears at first sight. Under certain assumptions (on what?), this indeed is true.

Polya Urn: Sample space?

Earlier, we had 'one shot' experiment; that is experiment which was not done sequentially, but done at one go. Here, in Polya urn scheme, we have experiment which is done sequentially. We have graduated to calculating probabilities without even thinking of the outcomes of the experiment, using only agreed-upon-rules.

We shall develop this attitude from now on. But if someone asks we should be able to tell what is the probability space (otherwise, we are working in the air).

Just to convince you that no cheating is going on, here is the sample space. Let us start with only one green ball and one red ball. let us imagine we have a supply of balls $(r_n : n \geq 1)$ and $(g_n : n \geq 1)$ and when needed we take these balls in that order. We start with urn $\{r_1, g_1\}$.

For one draw:
sample space is $\{r_1, g_1\}$; probabilities each 1/2.

For two draws:
sample space

$$\{r_1 r_1; \quad r_1, r_2, \quad r_1 g_1, \quad g_1 g_1, \quad g_1 g_2, \quad g_1 r_1\}$$

each probability 1/6.

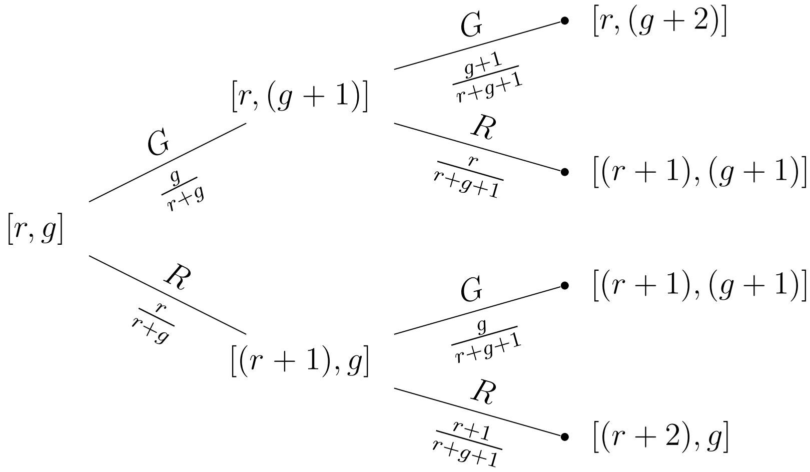
Note that there is no outcome r_1, g_2 or $g_1 r_2$ simply because when the first one is r_1 then there can not be g_2 next.

similarly for three draws also you can make. Thus the sample space keeps on changing. You can draw a tree diagram depicting the composition of the urn at each stage and the outcomes etc. However such a tree is not useful and I would not like to stress. We should understand probabilities rather than individual outcomes. Instead, tree diagram giving probabilities is more useful and this is what I shall do below starting with general (r, g) urn.

If you read the letters along any particular path, you will get events. Probability of the event is simply product of the numbers on the branches along the path.

On each branch is written conditional probability of that event given the past till then.

At the end of each branch is written the composition of the urn to help you calculate probabilities for the next branch. Initially composition is r red and g green. After green draw urn has become $[r, g+1]$ etc.



Friedman urn:

Urn has r red and g green balls.
Pick a ball at random; see colour; put it back;
add a ball of opposite colour. *Repeat.*

This is called Friedman urn model.

You can calculate:

$$P(R_1) = \frac{r}{r+g}; \quad P(G_1) = \frac{g}{r+g}$$

$$P(R_2) = \frac{r}{r+g}\frac{r}{r+g+1} + \frac{g}{r+g}\frac{r+1}{r+g+1}.$$

You can see that they are different, in general.

In this model 'there is a central tendency': if there are more red balls, you are likely to see red ball and add green ball; if there are more green balls you are more likely to see green ball and add red ball.

This models 'safety campaign'. Red balls are accidents and green are safety measures. If there are more accidents, police wake up and implement more and more safety measures. Once the safety measures increase and reduce accidents, police will be lax and accidents start increasing.

Several interesting calculations can be made with this model too. But we move on.

Multiplication rule:

In an experiment, the definition of conditional probability already told us

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1).$$

We can generalise this. For any n events $A_1, \cdots, A_n$

$$P(\bigcap_1^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n| \bigcap_1^{n-1} A_i).$$

this immediately follows by using the definition of conditional probability on the right side. This is called **Multiplication rule**. Just keep in mind that probability of intersection is not just product of probabilities of the events; but successive conditional probabilities.

Remember you can express intersection of events, in any order and so in the right side of the formula above, you can interchange the indices using any permutation. For example

$$P(\bigcap_1^n A_i) = P(A_n)P(A_{n-1}|A_n)P(A_{n-2}|A_n \cap A_{n-1}) \cdots P(A_1|\bigcap_2^n A_i).$$

Sampling without replacement:

Method 1: Suppose there are 20 items $S = \{1, 2, 3, \cdots, 20\}$. Sampling WOR means the sample space consisting all five tuples $(x_1, \cdots, x_5)$ of distinct elements from S. Each outcome has probability $1/(20)_5$. This is how we described earlier.

Method 2: We can consider the following sequential experiment like the urn models above: Pick a point from S at random; then pick a point at random from among the remaining 19; then pick a point at random from among the remaining 18; pick 5 times.

What are the outcomes? all possible five tuples $(x_1, \cdots, x_5)$ as above. What is probability of such an outcome? say $\omega = (x_1, \cdots, x_5)$. Use multiplication rule. $p(\omega)$ is a product of the following:

Probability that first choice is x_1; Probability second is x_2 give first is x_1 etc.

since first choice is: pick one at random; probability of picking x_1 is 1/20. Next is picked from among the remaining 19 at random. So chance of picking x_2 given first was x_1 equals 1/19 and so on. Thus, keeping multiplication multiplication in mind we attribute

$$p(\omega) = 20 \cdot 19 \cdot 18 \cdot 16 \cdot 15.$$

Thus in this second method we do sequentially and attribute probabilities keeping multiplication in mind.

Thus both outcomes and their probabilities are same as in Method 1.

Thus both methods can be regarded as sampling WOR. It is more convenient to think of doing the experiment in the second fashion, sequentially.

Stop a minute and think for a while. There is subtle difference between the two methods. Usually one says 'they are same', without knowing what this means. Superficially they are not same. First method is 'one shot' affair; picking a five tuple from among the sample points, equally likely. Second method is a sequential procedure, no five tuple is picked; at each stage only one number is picked.

What we did above is to show that both methods lead to the same outcome space with same probabilities. So from now on, you can think of sampling WOR as method 2 as well.

independence:

Let us pick a card at random from the deck of cards. Let A be the event that the card is Ace and B is the vent that the card is a spade. You can see

$$P(A) = \frac{4}{52} = \frac{1}{13}; \qquad P(A|B) = \frac{1/52}{13/52} = \frac{1}{13}.$$

in other words the conditional probability is same as the unconditional probability. The information that the card is spades did not influence the chances of being an Ace. When this happens it appears reasonable to call the events independent.

Thus events A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

More generally events $\{A_1, A_2, \cdots, A_n\}$ are **independent** if for any choice, $1 \leq i_1 < i_2 < \cdots < i_k$

$$P(\overset{k}{\underset{j-1}{\cap}} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j}).$$

Thus probability of any k of them simultaneously happening should equal product of probabilities of those k events.

If you have probabilities then definition of independence can be used to check whether certain events are independent. Also, if you are given some probabilities and told that certain events are to be independent, then you can assign probabilities confirming to these demands.

sampling with replacement:

Let us consider again $S = \{1, 2, \cdots, 20\}$

Method 1: consider all the tuples $(x_1, \cdots, x_5)$ where each $x_i \in S$. There are 20^5 such outcomes and they are equally likely.

Pick any five numbers $(a_i : 1 \leq i \leq 5)$ from S. You can show the following events are independent.
$A_i = i$-th entry of the five tuple is a_i; $\quad 1 \leq i \leq 5$.

Method 2: Pick a number at random from S. Repeat this experiment independently till five choices are made.
When you do this experiment you get a five tuple $\omega = (x_1, x_2, \cdots, x_5)$ with each $x_i \in S$. What probability should be assigned to this sample point? They said the experiment is picking a number at random from A. So picking x_i at i-th draw has chances $1/20$. They said that the choices are to be independent. This means that events regarding different choices should be independent events. In particular Chances that i-th choice is a_i for $1 \leq i \leq 5$ should be product of their individual probabilities. Thus, in particular probability of this outcome ω should be $1/20^5$ same as above.

In other words, in this method 2, we assigned the probabilities confirming to the specifications of independence.

However, both methods lead to the same sample space and same probabilities for the outcomes. So from now on sampling WR need not be regarded as one shot affair like selecting on five tuple; it can be regarded as sequential procedure.

Example:
Have a coin whose chance of heads is 0.3. Toss it independently twice.

If we toss it once the outcomes are $\{H, T\}$ with probabilities 0.3 and 0.7 respectively.

If we toss twice the outcomes are $\{HH, HT, TH, TT\}$. Repeating the earlier argument that probability of HT is product of probabilities (because they said questions regarding different tosses should be independent) Thus

$$p(HH) = (0.3)^2; \quad p(HT) = p(TH) = (0.3)(0.7); \quad p(TT) = (0.7)^2.$$

You can actually talk about 'repeating an experiment independently'. This and random variables is what we discuss next.