Fun with Balls and Coins

B V Rao

B V Rao was a Professor at the Indian Statistical Institute, Kolkata. Currently he is with the Chennai Mathematical Institute.

Even elementary probability has a lot of interesting things to offer to the students to get excited and think about. Unfortunately our curricula are generally designed in such a way that interesting things are carefully left out. Even if the material is touched upon briefly, the discussion is reduced to simple calculation of ratios devoid of any meaning. In this article we discuss two interesting models in probability. We just throw balls into boxes as dictated by tossing coins, nothing more than that.

Conditional Probabilities

Suppose that I am tossing a coin four times. What are the possible outcomes? Since each toss results either in Heads or Tails (we discount the possibility of the coin falling on its edge), there are 16 outcomes. You can write all the outcomes as:

HHHH, HHHT, HHTH, HHTT,
HTHH, HTHT, HTTH, HTTT,
THHH, THHT, THTH, THTT,
TTHH, TTHT, TTTH, TTTT.

If the coin is fair, then all the outcomes are equally likely, so we assign probability $1/16$ to each outcome. Suppose that the coin is not fair; then in a single toss Heads will come up with probability p, say, where $0 < p < 1$. Then for an outcome in which H appears k times we assign the probability $p^k(1-p)^{4-k}$. Thus for the outcome HHHH we assign the probability p^4 whereas for the outcome HTHT we assign the probability $p^2(1-p)^2$. The reason for assigning probabilities for outcomes in this way is, what is called, the assumption of independent tosses. We shall return to this point again soon.

Keywords
Bose–Einstein statistics, Chandrasekhar's model of diffusion, probability, conditional probability.

An event is any set of outcomes. For example, 'number of Heads is one' describes an event and it consists of the four outcomes

$$\{\mathrm{HTTT}, \mathrm{THTT}, \mathrm{TTHT}, \mathrm{TTTH}\}.$$

Probability of an event is just the sum of probabilities of the outcomes which are in that event. Thus probability of the event just described is $4p(1-p)^3$, because the event has four outcomes and each of the four outcomes has probability $p(1-p)^3$. By the very definition of probability of events, it is clear that if we have two events A and B which have no outcome in common then

$$P(A \cup B) = P(A) + P(B). \quad (1)$$

Of course, this holds for more than two events also, provided there are no outcomes in common between any two of the events. In fact, this holds even for an infinite sequence of events, $A_1, A_2, \cdots$, provided no two of the events have any common outcome; in other words, the events are pairwise disjoint. Though in the present experiment of tossing a coin we do not have infinitely many pairwise disjoint events, the fact that probabilities add up even for a sequence of disjoint events will be useful later.

Probability of an event is just the sum of probabilities of the outcomes which are in that event.

Let us note in passing that, in the above experiment, the chances of having exactly k Heads equals $\binom{4}{k}p^k(1-p)^{4-k}$ for $0 \le k \le 4$. More generally, if we perform n tosses of a coin whose chance of heads in a single toss is p, then the event A_k: 'the number of heads is exactly k', has probability $\binom{n}{k}p^k(1-p)^{n-k}$. This is again because any single outcome in which H appears k times and T appears $n-k$ times has probability $p^k(1-p)^{n-k}$ and there are $\binom{n}{k}$ many such outcomes. Of course, as mentioned before, we are making independent tosses. These probabilities are aptly called binomial probabilities.

If we perform n tosses of a coin whose chance of heads in a single toss is p, then the event A_k: 'the number of heads is exactly k', has probability $\binom{n}{k}p^k(1-p)^{n-k}$. These probabilities are aptly called binomial probabilities.

Let us now turn to another useful concept, namely that of conditional probability. In practice, we perform an experiment and possess some partial information about the outcome. For example, I toss a fair coin 4 times and I tell you that at least three heads appeared. Given this information, what are the chances that four heads appeared. Let B be the event 'at least three heads appeared' and A be the event 'all four are heads'. Thus our interest is in the conditional chances of A given B.

To avoid confusion with $P(A)$, let us denote this conditional probability by $P(A|B)$. Here $B = \{$HHHT, HHTH, HTHH, THHH, HHHH$\}$ and A consists of one outcome, namely, HHHH. Having known that one of the five outcomes in B appeared, the chances of the outcome HHHH appearing would naturally be $1/5$, if the coin is fair (since then all the outcomes are equally likely). On the other hand, if the coin has chance of heads p, instead of counting the number of outcomes we shall count their probabilities. The total probability of the event B is $p^4 + 4p^3(1-p)$ out of which A carries probability p^4 and thus it is natural to declare $P(A|B) = p^4/[p^4 + 4p^3(1-p)]$. If B is the same event as above, but C is the event 'exactly one Head appeared', then it is only natural to feel that given B, the event C could not have occurred and hence its conditional probability is zero.

For an outcome ω, let us denote by p_ω the probability of the outcome ω. With this notation, if $A \subset B$, then $P(A|B) = [\sum_{\omega\in A} p_\omega]/[\sum_{\omega\in B} p_\omega] = P(A)/P(B)$, whereas, if $A \cap B = \emptyset$, then $P(A|B) = 0$. We can combine both these into one single equation as $P(A|B) = P(A\cap B)/P(B)$. In case $A \subset B$, or in case $A \cap B = \emptyset$, this formula reduces to the earlier formulae. As mentioned earlier, the reason for the appearance of intersection of A and B is the following. We are informed that an outcome from the event B occurred and we want the chances that an outcome from A occurred – well, this is the same as asking the chances that an outcome from $A\cap B$ occurred, because we already know that an outcome outside B could not have occurred, with the information at hand. There are other reasons for adopting this definition of conditional probability, but we do not wish to dwell on this matter too long.

One upshot of this definition is

$$P(A \cap B) = P(B)P(A|B). \qquad (2)$$

Of course, just like (1), this can also be stated for more than two events. For example, for three events A, B and C;

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

This is so because, by our definition, $P(B|C) = P(B \cap C)/P(C)$ and $P(A|B \cap C) = P(A \cap B \cap C)/P(B \cap C)$. This is a very intuitive and pleasing formula. If you want all three events to happen, then first of all the event C must occur; secondly, given this information the event B should occur; and finally given that these two events occurred, the event A should also occur. Mathematically, it makes no difference how you order the events. For example, you can say: first A should occur; and given that A occurred the event B should occur and given that these two events occurred the event C also must occur. To put it mathematically, the equation $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$ is also true. Which formula to use is dictated from what we know and what we do not know, that is, the information we have decides it. You will see later.

Let us note that the conditional probability $P(A|B)$ as a function of A, for fixed B, also obeys the same rules as probability does. For example, addition rule holds, that is, if A_i for $i \geq 1$ are disjoint events, then $P(\cup_i A_i|B) = \sum_i P(A_i|B)$. This is immediate from the definition of conditional probability and the observed rule for probability.

The conditional probability $P(A|B)$ as a function of A, for fixed B, also obeys the same rules as probability does.

One interesting situation is when the conditional probability $P(A|B)$ is the same as the unconditional probability $P(A)$. Thus the information 'B occurred' did not influence us to change the value of the probability of A. In such a situation, we say that A and B are independent events.

Independent Events

One interesting situation is when the conditional probability $P(A|B)$ is the same as the unconditional probability $P(A)$, that is, $P(A \cap B) = P(A)P(B)$. Thus the information 'B occurred' did not influence us to change the value of the probability of A. In such a situation, we say that A and B are independent events. For example, in the above experiment of tossing a coin four times, let A be the event: 'first toss results in Heads' and B be the event: 'fourth toss results in Heads'. Then A consists of the eight outcomes which have H in the first place; B consists of the eight outcomes which have H at the last place and $A \cap B$ consists of the four outcomes that have H at the first and last place. If you calculate, you will see that $P(A) = p$; $P(B) = p$, $P(A \cap B) = p^2$. Thus $P(A|B) = p^2/p = p = P(A)$. Thus the events A and B are independent. Perhaps common sense has already guided you to this conclusion.

More generally, events $A_1, A_2, \cdots A_n$ are said to be independent if the following holds: whenever you take an integer k with $1 \leq k \leq n$ and take any k among those n events, then probability that all these k events occur simultaneously is the same as the product of their individual probabilities. I have stated in words, but you should be able to write in symbols what this means. The reason for adopting this definition is the following. Suppose you compute the chances of 'A_1 and A_3 occuring and A_8 not occuring'. That is $P(A_1 \cap A_3 \cap A_8^c)$. We use B^c to denote complement of the event B. Someone told you that A_2 occured, but A_6 and A_7 did not. Then will the probability of the earlier event change? It does not. That is, we have

$$P(A_1 \cap A_3 \cap A_8^c) = P(A_1 \cap A_3 \cap A_8^c \,|A_2 \cap A_6^c \cap A_7^c) \ .$$

In fact, it is possible to formulate and show precisely that none of the events influence the others, in the sense of conditional chances being same as the unconditional chances. Conversely if we demand such carefully formulated equations to hold then we are led to the definition above.

Thus when we associated probabilities in the experiment of tossing a coin four times, we assumed that the tosses are independent. This is a very interesting situation. To start with, we do not have probabilities for the outcomes and hence we do not have probabilities for events as well. If we had, then it makes sense to calculate probabilities of certain events and verify whether they are independent or not. We are told that the tosses have to be independent and we assigned probabilities to the outcomes so that the hypothesis of independence holds. For example, if for $1 \leq i \leq 4$, A_i is the event: 'i-th toss results in Heads'; then independence of these four events forces you to assign probability p^4 for the outcome HHHH. This is what made us assign probabilities, the way we did, for the outcomes.

Bose–Einstein Urn

We shall now discuss a simple model in probability. I have two urns, urn I and urn II. I have four balls and I want to distribute these balls into the two urns. It is enough to keep track of the number of balls in urn I, because the remaining number should be in urn II. Here is a way to distribute the balls into the urns. Take a ball, toss your coin and put the ball in urn 1 if Heads up, otherwise put in urn II. Repeat this for all the four balls, thus tossing the coin once for each ball to decide where it goes. The chances of 'exactly one ball in Urn I' is same as the chances of getting exactly one Head. So from what we discussed above, this equals $4p(1-p)^3$.

Suppose that I have three coins with chances of heads in a single toss being p_1, p_2 and p_3. I pick one coin at random and I use that coin to distribute all the four balls by the above rule. Then what are the chances that urn I has exactly one ball? We shall put formulae

Thought experiments are the best way to understand phenomena and also to have fun beyond the technicalities.

(1) and (2) to work. The event A is 'urn I has exactly one ball'. Define B_i to be the event 'coin p_i is chosen'. Our event A is the union of the three disjoint events $A \cap B_1$, $A \cap B_2$ and $A \cap B_3$ – because we must have chosen exactly one of the three coins. Consequently, $P(A) = \sum P(A \cap B_i)$ by formula (1). Since $P(B_i) = 1/3$ and $P(A|B_i) = 4p_i(1-p_i)^3$, we now get by application of formula (2)

$$P(A) = \frac{4p_1(1-p_1)^3}{3} + \frac{4p_2(1-p_2)^3}{3} + \frac{4p_3(1-p_3)^3}{3}.$$

Suppose now, I have a bag full of coins, one for each p, $0 < p < 1$. It is understood that coin p has chance of Heads p in a single toss. I pick a coin at random and having picked the coin, I use it to distribute all the four balls into the urns, as earlier. We wish to find the chances of the event A: 'exactly one ball in urn I'.

This may appear to you as a meaningless question. You must be wondering if I could get so many coins at all! You are right, I cannot (No one can!). Actually we formulated the experiment in a picturesque way. We do not need so many coins. We could as well pick a number at random from the interval $(0, 1)$ and just take a coin whose chance of heads in a single toss is that p.

In any case, this is a thought experiment and the question itself is a very meaningful one – and as you will see later, it is useful too. You should know that thought experiments are the best way to understand phenomena and also to have fun beyond the technicalities. For fun (following Einstein), imagine that two photons are traveling parallel; I sit on one of them and observe the other, then should it not look stationary to me (since all photons travel at same speed)? For fun, imagine Ishant Sharma taking a bowling run of 160,000 km/sec and releasing the ball with an initial speed of 160,000 km/sec. Would the ball have a speed of 320,000 km/sec?

Returning to our problem, if you extend your imagination, the chances of our event above will again be an average, as earlier. The only difference is that, now it is a continuous average, simply because I have tooooo many coins. Such continuous averages are nothing but what we call integrals. Thus

$$P(A) = \int_0^1 4p(1-p)^3 \mathrm{d}p = \frac{1}{5}.$$

In fact you will realize that if A_i is the event: 'exactly i balls in urn I', then $P(A_i) = 1/5$ for $0 \leq i \leq 4$. You only need to do integration by parts.

Thus our calculations yield that the five events have equal probability, namely, 1/5. Let me explain this phenomenon in a different way. Assume that all the balls look alike. Then how many outcomes of the experiment can our eye distinguish? After all, since the balls look alike, it makes no difference to our eye whether you put first ball in urn I and the rest in urn II or you put second ball in urn I and the rest in urn II. In fact, there is nothing to distinguish first ball from the second. Our eye can only distinguish the difference in the outcomes by the number of balls in urn I. Clearly there are five possibilities, namely, urn I has i balls: $0 \leq i \leq 4$. Postulate (that is, make an axiom) that *the distinguishable outcomes are all equally likely*. You end up saying that the chances of having any one of these possibilities is 1/5. Thus we arrive at the same answer as above. We did not now use the thought experiment. Instead, we made two postulates: (1) all the balls look alike and (2) the distinguishable outcomes are equally likely.

Well, imagine for a moment that the balls are not really balls, they are photons. Imagine that the urns are not really urns, but two energy levels: 1 and 2. I am not distributing four balls, but the four photons are organizing themselves into the two available energy levels. How do they distribute themselves? The answer is: for

$0 \le i \le 4$, the chances of i photons having energy level 1 is $1/5$. Thus you have before you the discovery of the famous physicist Satyendranath Bose[1]. He was explaining Planck's formula for the distribution of energy in the radiation of a black body. Though this law is quantum mechanical in nature, all its derivations at that time were based on classical physics, an unhappy situation. Bose proposed this postulate in his paper 'Planck's Law and the Light quantum Hypothesis' (*Zeitschrift fur Physik*, 1924)[2].

[1] See *Resonance*, Vol.1, No.2, 1996.

[2] My source is the book *Satyendranath Bose: The man and his work* published in 1994 by S N Bose National Centre for Basic Sciences, Calcutta.

The interpretation in terms of the thought experiment is due to the statistician Sudhakar Kunte (from the journal *Sankhya* (1977)). Initially Bose proposed in terms of the two axioms (1) and (2) mentioned above. To understand the courage needed to put forward such a proposal consider the following. I have an urn with 100 balls; 1 red and 99 black. All black balls look alike. I draw a ball at random. What are the chances that the ball drawn is black? Can I pretend that there are only two distinguishable outcomes (since all black balls look alike) and say each has chance $1/2$?

Having got the basic idea, you can now make some embellishments, purely technical in nature (and easy). What happens if we still have two urns but n photons, instead of four? The answer is: urn I has k balls with probability $1/(n+1)$ for $0 \le k \le n$. This is easy to see; you only need to calculate the integral,

$$\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \mathrm{d}p$$

to arrive at the answer.

Moreover what if we have r energy levels instead of two? The chances of having a particular occupancy of the photons in the various energy levels is $1/\binom{n+r-1}{r-1}$. This can be derived following any one of the above methods. If you go by the thought experiment, then you should pick a die at random from all possible dice with r faces. More precisely, suppose that you have a box full of r-faced dice, one for each $(r-1)$-tuple $(p_1, \cdots, p_{r-1})$, where each $p_i > 0$ and $\sum p_i < 1$. This die when rolled will land up on face i with probability p_i for $1 \leq i \leq r$, where $p_r = 1 - \sum_{1}^{r-1} p_i$. Each face corresponds to an energy level. Select a die at random and roll the selected die n times to distribute all the n balls. Instead of binomial probabilities, one has to consider, what are known as, multinomial probabilities and instead of one integral one has to consider a multiple integral. But there is no new idea. To understand the calculations involved, we shall illustrate the case $r = 3$.

Thus we have n photons trying to organize themselves in three energy levels: 1, 2, and 3. Consider the region

$$\Delta = \{(p_1, p_2) : p_1 > 0; p_2 > 0; p_1 + p_2 < 1\}.$$

The photons pick a point at random from this region, say, (p_1, p_2). They pick a three-faced die. In a single throw, this die comes up face 1, 2 and 3 with chances p_1, p_2, and $p_3 = 1 - p_1 - p_2$ respectively. Each photon tosses the die and occupies the energy level showed by the die. Let us fix three non-negative integers n_1, n_2, n_3 such that $n_1 + n_2 + n_3 = n$ and ask the question: what are the chances that there are n_i photons in the energy level i? Equivalently, what are the chances of the event

$$A = \{\text{face } i \text{ occurs } n_i \text{ times, for } i = 1, 2, 3\}.$$

Let us understand the experiment. Now each throw can result in any one of the three faces 1, 2 and 3. Hence the n throws will result in 3^n outcomes, namely, all possible sequences of length n consisting of the digits 1, 2 and 3. If we take one such outcome $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ in which the digit i appears n_i times, then the probability of the outcome ω is given by $p_1^{n_1} p_2^{n_2} p_3^{n_3}$ (remember independent throws). But the number of outcomes in which the digit i appears n_i times for $i = 1, 2, 3$ is

$$\frac{n!}{n_1!n_2!n_3!}.$$

If B denotes the event '(p_1, p_2) is chosen from Δ', then the conditional probability $P(A|B)$ is given by

$$\frac{n!}{n_1!n_2!n_3!}p_1^{n_1}p_2^{n_2}(1-p_1-p_2)^{n_3} = f(p_1, p_2), \quad \text{say.}$$

As earlier, to get the unconditional probability of the event A we need to take the average over Δ. Unlike in the earlier case when the interval $(0, 1)$ has length one, now the area of the region Δ, which is a triangle, equals $1/2$. Just as average of certain numbers is their sum divided by the number of summands, average of a function over a region is the integral over the region divided by the area of the region. Thus the probability of the event A is

$$2\int\int_{0<p_1,p_2,p_1+p_2<1} f(p_1, p_2,)\mathrm{d}p_1\mathrm{d}p_2 \ .$$

Fix p_1 and integrate w.r.t. p_2 from 0 to $1 - p_1$ by substituting $x = (1 - p_1)p_2$ so that $\mathrm{d}x = (1 - p_1)\mathrm{d}p_2$ and now x ranges from 0 to 1. Then integrate w.r.t. p_1. If you do correctly, you will see that the integral does not depend on n_1, n_2, n_3 and equals $\binom{n+1}{2}$.

If you go by the 'balls looking alike' postulate, you need to count the number of outcomes that your eye can distinguish and you will see that this is a counting problem leading to the answer $\binom{n+r-1}{r-1}$. These probabilities go by the name of Bose–Einstein statistics. Following Feller, here is a quick justification for the answer mentioned. Put $(n + r - 1)$ star marks in a row. Select $(r - 1)$ of these star marks and make them vertical lines. Think of these $(r - 1)$ lines as walls so that you have before you r urns: urn 1 is the space upto the first line, urn r is the space after the last vertical line and the spaces between successive vertical lines give urns $2, 3, \cdots, (r-1)$. The number of star marks remaining, after converting the $(r-1)$ stars as vertical lines, is n and think of them as balls. This picture gives a possible distribution of the balls in the urns. The main points that need checking, which is easy, are (i) different choices of the star marks lead to distinct distribution of balls and (ii) every distribution of the balls can be obtained by suitable choice of the star marks. Clearly we can choose $(r-1)$ star marks out of the $(n+r-1)$ star marks in $\binom{n+r-1}{r-1}$ ways.

S N Bose's work on particle statistics which clarified the behaviour of photons and opened the door to new ideas on statistics of microsystems that obey the rules of quantum theory, was one of the top ten achievements of 20th century Indian science.

Thus to put matters picturesquely, if a certain number of photons wish to organize themselves among various available energy levels, they collectively pick up at random, a die with as many faces as there are energy levels and then each one obeys what the die tells it to do!

As remarked by the astrophysicist Jayant Narlikar: S N Bose's work on particle statistics which clarified the behaviour of photons and opened the door to new ideas on statistics of microsystems that obey the rules of quantum theory, was one of the top ten achievements of 20th century Indian science and could be considered in the Nobel Prize class. (See the Wikipedia article on S N Bose.)

Before moving on, let us note an amusing fact. It appears, Bose was trying to explain why the theory (at that time) was inadequate to explain the experimental evidence. He made an error in calculations and ended up adequately explaining the experimental evidence! So dear students, do not worry if you make mistakes, but learn from them.

Chandrasekhar Urn

In the model discussed above, we just distributed the balls into the urn and there the matter ended. In a sense, it is static – unless, you want to change the parameters and try to understand appropriate limits, when

We shall now discuss a dynamic model, that is, the number of balls in the urn evolves with time. It is really such models that are regarded as urn models in probability.

such limits exist. Even then, the model itself is static, you will be doing the same experiment with different parameters, that is all. Actually, such static models are not considered as true urn models in probability. Such models are aptly called 'occupancy models'. We shall now discuss a dynamic model, that is, the number of balls in the urn evolves with time. It is really such models that are regarded as urn models in probability.

I need a little more of your patience in understanding this model. Now I consider only one urn. I have an unlimited supply of balls with me. Every morning, I look at the urn, remove some of the existing balls from the urn and then add some balls to the urn. I repeat this every morning. After several days how would the composition of the urn look like?

Well, we should make our question clear: What do we mean by the composition of the urn? We should also make clear the mechanism of removing and adding balls. I have a coin whose chance of heads in a toss is p, where $0 < p < 1$. I also have a number $\lambda > 0$. Using this data, we describe the mechanism.

Suppose that I see j balls this morning. I take a ball and toss my coin. Heads up, I remove the ball, Tails up, I return the ball back to the urn. I do this for each ball in the urn. Since the chance of keeping a ball is $(1-p)$, the chances that i balls remain after doing this process is $\binom{j}{i}(1-p)^i p^{j-i}$. These are the binomial probabilities mentioned earlier.

I shall add r balls with probability $e^{-\lambda}\lambda^r/r!$, where r could take any non-negative integer value. These are the so-called Poisson probabilities, which arise as appropriate limits of binomial probabilities. Thus the chances that I add zero balls equals $e^{-\lambda}$; the chances that I add 29 balls is $e^{-\lambda}\lambda^{29}/(29!)$. Even though, there is no *a priori* limit on the number of balls that I add, please do keep in mind that I add only a finite number of balls.

Thus the number of balls in the urn on any day is a finite number, perhaps zero, but never infinity.

The mechanism is now made clear. The question is made precise as follows. For an integer $k \geq 0$, denote by $p_k^{(n)}$ the probability of having k balls in the urn on day n. Of course, this probability depends also on how we started the whole game. But this dependence is not shown in the notation. The question is to see if the limit $\lim_n p_k^{(n)}$ exists and find it when it exists. A mathematician talks of limits, but in practice what it means is the following. Suppose the above limit is α_k. We can believe that on the millionth day the chances of finding k balls in the urn is α_k. (A slight warning: a sequence may converge to 0.9 but its first billion terms may be zero!).

Let us start with the simplest initial condition, namely, on day 0, there are no balls in the urn. This does not make the problem any uninteresting. The balls in the urn on day 1 are just the balls we have put into the urn. Thus, by the specification of the mechanism it is clear that

$$p_k^{(1)} = e^{-\lambda}\lambda^k/k!.$$

Now let us calculate $p_k^{(2)}$. This is nothing but the probability of the event A: 'there are k balls on day 2'. Unfortunately, according to our mechanism, to know probability of the event A, we should know how many balls are there in the urn on day 1. So, to pave the way for calculation, let us define, for integer $j \geq 0$, the event B_j: 'there are j balls in the urn on day 1'. Clearly the events $(B_j; j \geq 0)$ are disjoint and hence so are the events $(A \cap B_j; j \geq 0)$. Moreover their union is A: after all on day 1, the number of balls has to be something or the other. Thus by formula (1), $P(A)$ is just the sum of the numbers $P(A \cap B_j)$. The fact that we have infinitely many events should be of no concern. We already know that formula (1) is true for any sequence of disjoint events.

Now formula (2) will come into play. We know $P(B_j)$ already, calculated above. Thus we need to now calculate $P(A|B_j)$. So let us assume that there are j balls on day 1. How can we end up with k balls on day 2. We should keep certain number, say i, of balls and add $k-i$ balls so that the total is k, as desired. You cannot keep more than k balls because then no matter how many you add, you will never end up with k balls finally. Thus the number i should not exceed k. Of course i cannot exceed j; you cannot keep more than what it already has! Thus, denoting by $j \wedge k$ the minimum of the two integers j and k, we have $i \leq j \wedge k$. Next, we should put exactly $k-i$ balls so that we end up with exactly k balls. Let us define for $0 \leq i \leq k \wedge j$, the event A_i: 'i balls are kept and $k-i$ are added'. By independence of the mechanism of removing and adding balls, we get

$$P(A_i|B_j) = \binom{j}{i} p^{j-i}(1-p)^i \ \times \ e^{-\lambda}\lambda^{k-i}\frac{1}{(k-i)!}.$$

The event A is the union of the disjoint events A_i. Hence, by the addition rule for conditional probabilities, we get

$$P(A|B_j) = \sum_{i=0}^{k\wedge j} \binom{j}{i} p^{j-i}(1-p)^i e^{-\lambda}\lambda^{k-i}\frac{1}{(k-i)!}.$$

Thus

$$p_k^{(2)} = P(A) = \sum_{j\geq 0} P(B_j)P(A|B_j)$$

$$= \sum_{j\geq 0} e^{-\lambda}\frac{\lambda^j}{j!} \ \sum_{i=0}^{k\wedge j} \binom{j}{i} p^{j-i}(1-p)^i e^{-\lambda}\frac{\lambda^{k-i}}{(k-i)!}.$$

$$= e^{-\lambda}e^{-\lambda}\lambda^k \sum_{i=0}^{k} \frac{(1-p)^i}{(k-i)!i!} \sum_{j\geq i} \frac{\lambda^{j-i}p^{j-i}}{(j-i)!}$$

$$= e^{-[\lambda(1+q)]} \ \frac{[\lambda(1+q)]^k}{k!},$$

where we denoted $(1-p)$ by q. The last equality above is an exercise on interchanging the order of summation.

Exactly similar calculation, without needing any new ideas, will show that

$$p_k^{(3)} = e^{-[\lambda(1+q+q^2)]} \; \frac{[\lambda(1+q+q^2)]^k}{k},$$

and more generally for any integer $n \geq 0$,

$$p_k^{(n+1)} = e^{-[\lambda(1+q+\cdots+q^n)]} \; \frac{[\lambda(1+q+\cdots+q^n)]^k}{k!}.$$

Noting that $1+q+q^2+\cdots = 1/(1-q) = 1/p$, we get what we have been looking for.

$$\lim_n p_k^{(n)} = e^{-\lambda/p} \; [\lambda/p]^k \; \frac{1}{k!}.$$

Of course, all this calculation is done with the condition that there were zero balls in the urn to start with (on day zero). Actually the answer does not depend on the initial condition. Whatever be the initial number of balls in the urn, the limit of $p_k^{(n)}$ exists and is the same as above.

Well students, now imagine the following. Actually neither do I have a bag of balls with me nor am I removing and adding balls. Instead, I have a solution containing Brownian particles (under diffusive equilibrium, but ignore this phrase). I do not have an urn, but a geometrically well-defined volume V in the solution. During each time period, certain particles emerge from this volume V and certain particles will have entered this volume V. Time period is not one day, but is of the order of few hundredths of a second.

What you have before you is a model by the astrophysicist Subrahmanyan Chandrasekhar[3] (Nobel Prize in 1983) discussed in Chapter III of his paper 'Stochastic problems in Physics and Astronomy' in *Reviews of Modern Physics* (1943)[4].

Chandrasekhar was explaining the inner relationships that exist between the phenomena of Brownian motion,

[3]See *Resonance*, Vol.2, No.4, 1997.

[4] My source is: *Selected papers on noise and stochastic processes* edited by Wax Nelson, 1954, a Dover Publication.

diffusion, and fluctuations in molecular concentrations. Explicit calculations were needed to compare theory with experimental evidence. It was observed that the density fluctuations studied in terms of microscopic analysis of the stochastic motion of the individual particles are in complete agreement with the macroscopic theory of diffusion. The number p is termed the 'probability after-effect factor' – it is the probability of a particle leaving the volume.

As remarked by the astronomer R J Tayler: Chandrasekhar was a classical applied mathematician whose research was primarily applied in astronomy and whose like will probably never be seen again. (See the Wikipedia article on S Chandrasekhar).

Here is a brief reason for Poisson probabilities entering the model. Our focus is on the number of particles in a small volume. There are a large number of particles outside this volume and each has a small probability of entering this volume. With a good mathematical reasoning, such probabilities are modelled by Poisson probabilities.

What Next

We discussed two models. My idea was to impress upon you that ideas are simple and not difficult; applications are very concrete and not abstract; calculations are interesting and not boring; and such models as those we discussed, make learning fun and enjoyable. But of course, you would not like to stop with simple calculations. You would try to push things as far as they can go, or apply the ideas elsewhere and so on. While undertaking such a program, more mathematics comes into play. One of the best books to start learning these and other matters is that of William Feller (Croatian–American mathematician; July 7, 1906–January 14, 1970).

Suggested Reading

[1] W Feller, *Introduction to the theory of probability and its applications*, John Wiley, USA, Vol.1, 2000.

[2] Wikipedia articles on S N Bose and S Chandrasekhar.

[3] Paul Davies, *Super Force*, Simon & Schuster, USA, 1983.

Address for Correspondence
B V Rao
Chennai Mathematical Institute
Plot H1, SIPCOT IT Park, Siruseri, Padur Post
Chennai 603 103, TN, India.
Email: bhamidivrao@gmail.com