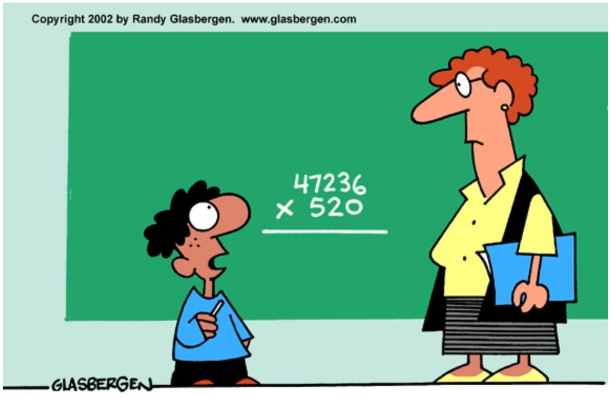

"AREN'T THERE ENOUGH PROBLEMS IN THE WORLD ALREADY?"

Conditional distributions and Conditional expectations:

You now have all the basic material: probability and conditional probability; random variables; expectation and variance and moments; distribution of random variables.

What comes next is combination of conditional probability with other concepts leading to conditional distribution, conditional expectation. There is an interesting and useful formula hidden in this development.

Consider the experiment: Take sample of size two with replacement from $\{0, 1, 2\}$. Let X be the maximum of the sample and Y is the second chosen point. Sample space is

$$\Omega = \{00;\ 01;\ 02;\ 10;\ 11;\ 12;\ 20;\ 21;\ 22\}$$

Each outcome has probability $1/9$.

The joint distribution of (X, Y) is given below.

■ *Warning:* Earlier I used X horizontal and Y vertical. Now to read values of X read vertically the first column and values of Y are on the top row.

We shall follow this notation. earlier when you read an entry in the matrix of probabilities a_{ij} was the probability $P(Y = i; X = j)$. I did not like it. Now it is $P(X = i; Y = j)$; sounds better.

In any case this change should not lead to confusion because table explains irrespective of any conventions. ∎

$X\backslash Y$	0	1	2	total
0	1/9	0	0	1/9
1	1/9	2/9	0	3/9
2	1/9	1/9	3/9	5/9
total	3/9	3/9	3/9	1

Sometimes we have partial information. For example you have info that $X = 2$. You would like to know the distribution of Y. Of course Y takes values $0, 1, 2$ and and against each value we should now list the conditional probabilities are $P(Y = 0|X = 2); (Y = 1|X = 2); P(Y = 2|X = 2)$. Thus conditional distributions of Y are as follows:

Given $X = 0$:

values of Y:	0	1	2
Conditional probabilities:	1	0	0

Given $X = 1$:

values of Y:	0	1	2
Conditional probabilities:	1/3	2/3	0

Given $X = 2$:

values of Y:	0	1	2
Conditional probabilities:	1/5	1/5	3/5

These are called conditional distributions of Y for the values of X given as shown. Note that these conditional probabilities do add to one in each case. You can calculate conditional expectations which means expectations under these conditional distributions. Thus $E(Y|X = 0)$, conditional expectation of Y given $X = 0$ is 0.

Similarly $E(Y|X = 1) = 2/3$ and $E(Y|X = 2) = 7/5$.

Remember distribution of Y is only one; whereas conditional distributions are many; one for each given value of X. Similarly expectation is just a number where as conditional expectations of Y are many numbers; one for each given value of X.

We want to summarize all these conditional expectations into one quantity; not a number but a function. It will be defined on the sample space Ω as follows. Take any sample point ω suppose $X(\omega) = a$ then we define $Z(\omega) = E(Y|X = a)$. Thus for all sample points ω for which $X(\omega) = a$ we have the same value for the function Z. This Z is called conditional expectation of Y given X, denoted $E(Y|X)$.

So keep in mind $E(Y|X)$ is not a number; it is a function on the sample space. There is again no confusion because we only said 'expectation of Y given X' and did not say any specific value of X. Hence it must encode all the information and hence it is the function Z rather than any one of the earlier numbers.

Being a function on the sample space, it is a random variable. So it makes sense to take its expectation once again!

Let us quickly do all these considerations in generality.

Suppose (Ω, p) is a probability space and X, Y are random variables. Suppose their joint distribution, given in a bivariate table is as follows.

X takes values $\{x_i : i = 1, 2, \cdots\}$ may be finite or infinite set.
Y takes values $\{y_j : j = 1, 2, \cdots\}$ may be finite or infinite set.

$$p_{ij} = P(X = x_i, Y = y_j)$$

The so called marginal totals are the following.

$$p_{i\bullet} = \sum_j p_{ij}; \quad p_{\bullet j} = \sum_i p_{ij}; \quad i, j \geq 1$$

Thus the distribution of X and Y are given by

$$P(X = x_i) = p_{i\bullet} : \quad i = 1, 2, \cdots$$

$$P(Y = y_j) = p_{\bullet j} : \quad j = 1, 2, \cdots .$$

Now we can define the conditional distribution of Y given $X = x_i$ is the following:

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\bullet}} : \quad j = 1, 2, \cdots .$$

this is defined for each $i = 1, 2, \cdots$ and whatever be i all the numbers above add to one when summed over j.

$$E(Y|X = x_i) = \sum_j y_j \frac{p_{ij}}{p_{i\bullet}}$$

Finally $E(Y|X)$ is the following function:

$$E(Y|X)\ (\omega) = E(Y|X = a); \quad \text{where} \quad a = X(\omega)$$

Thus the function $E(Y|X)$ takes same value at two sample points ω and η if $X(\omega) = X(\eta)$. As a consequence it is a function of X, namely

$$E(Y|X) = \varphi o X; \quad \text{where} \quad \varphi(x_i) = E(Y|X = x_i)$$

Of course we have not defined the function φ at real numbers other than the x_i s. But if you are worried about that put $\varphi(x) = 0$ if x is not any of the numbers $\{x_i\}$. These do not matter. After all when when you compose two functions $f(g(x))$, values of f on points in the range of g matter. Think about it.

As a consequence of the change of variable rule

$$E[E(Y|X)] = \sum_i \varphi(x_i) P(X = x_i)$$

$$= \sum_i \sum_j y_j \frac{p_{ij}}{p_{i\bullet}} p_{i\bullet} = \sum_j \sum_i y_j p_{ij} = \sum_j y_j p_{\bullet j}$$

$$\sum_j y_j P(Y = y_j) = E(Y)$$

Thus we have proved

$$\blacksquare \qquad E[E(Y|X)] = E(Y) \qquad \blacksquare$$

This is a very useful formula. It allows you to calculate expected values in complicated situations. here is an example.

Example:

Imagine there is a maze with four doors:
T, R, B, L : Top/Right/Bottom/Left.

If rat exits through T, it is immediately out. If exits through B, then it travels for 2 minutes and then it is out. If exits through R, it travels for 4 minutes and then returns to the starting place. If exits through L, then returns to the starting place after 5 minutes.

Assume that the rat always chooses one of the four doors at random. What is the expected time to exit from the maze?

You can write down the sample points, for each sample point calculate its probability $p(\omega)$ and the time $Y(\omega)$ to exit for that sample point ω and calculate $\sum Y(\omega)p(\omega)$.

Here is a simpler way. Let Y be the time to exit and X be the first choice of Door. Suppose $E(Y) = a$. Clearly

$$E(Y|X = T) = 0; \qquad E(Y|X = D) = 2.$$

$$E(Y|X = R) = 4 + a$$

because the conditional distribution of Y given $X = R$ is same as the distribution of $4 + Y$. similarly

$$E(Y|X = L) = 5 + a.$$

Thus using the earlier formula,

$$a = E(Y) = E[E(Y|X)] =$$

$$\left[0 \times \frac{1}{4}\right] + \left[2 \times \frac{1}{4}\right] + \left[(4 + a) \times \frac{1}{4}\right] + \left[(5 + a) \times \frac{1}{4}\right].$$

or

$$a = \frac{11}{2}$$

Best Fits:

I have a sample space (Ω, p) and two random variables X and Y. Assume that they have finite second moment.

We wish to estimate Y. Here are specific questions.

Question 1: What is the best constant fit to Y?

Question 2: What is the best 'linear in X' fit to Y?

Question 3: What is best 'function of X' fit to Y?

Q1:

Meaning of the question is the following. What is the best number a so that $E(Y-a)^2$ is the least possible. If your guess is the number a and if then one observes Y; you will incur a loss $(Y-a)^2$. Which number a minimizes your loss? This loss is called squared error loss. $(Y-a)$ is the error.

If μ and σ^2 are the mean and variance of Y, then

$$E(Y-a)^2 = E(Y-\mu)^2 + E(\mu-a)^2 + 2E\{(Y-\mu)(\mu-a)\}$$
$$= \sigma^2 + (\mu-a)^2$$

So best guess and expected penalty for this guess are

$$\mu_Y; \qquad \sigma_Y^2 \qquad (\bullet)$$

Q2:

Again, 'best' is interpreted as earlier. you have to prescribe a linear function of X, say, $a+bX$ (if you are fussy and do not want to use the word linear for $a+bX$, agreed, use the word affine). Of course you know the joint distribution of (X,Y).

Now we make observation (x,y). Thus observed value of Y is y and your guess is $a+bx$. The penalty is $(y-a-bx)^2$. We must minimize penalty, makes no sense because it random. So minimize expected penalty $E(Y-a-bX)^2$. Since we want to minimize over (a,b) we set gradient equal to zero. Interchanging differentiation and expectation, the equations are

$$2E(Y-a-bX)=0; \quad \text{or} \quad \mu_Y = a + b\mu_X.$$

$$2E\{X[Y-a-bX]\}=0 \quad \text{or} \quad E(XY) - a\mu_X - bE(X^2) = 0$$

Substituting for a from the first equation we get

$$E(XY) - \mu_X[\mu_Y - b\mu_X] - bE(X^2) = 0$$

$$\sigma_{XY} - b\sigma_X^2 = 0;$$

$$b = \frac{\sigma_{XY}}{\sigma_X^2}; \qquad a = \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X.$$

Strictly speaking you need to find second derivative to decide whether this is max or min or none! let us skip that part.

What is the expected penalty for such a guess?

$$E\left\{(Y-\mu_Y)-\frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X)\right\}^2$$

$$=\sigma_Y^2+\frac{\sigma_{XY}^2}{\sigma_X^2}-2\frac{\sigma_{XY}}{\sigma_X^2}\sigma_{XY}=\sigma_Y^2-\frac{\sigma_{XY}^2}{\sigma_X^2}$$

Remembering that $\sigma_{XY}=\rho\sigma_X\sigma_Y$, the above equals

$$\sigma_Y^2[1-\rho^2]$$

Thus the best guess and expected penalty for this guess are

$$\mu_Y+\frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X); \qquad \sigma_Y^2[1-\rho^2] \quad (\bullet\bullet)$$

comparing with $(\bullet)$ and remembering that $|\rho|\leq 1$ you see that this guess using X reduces the penalty and the percentage reduction is in terns of the correlation between the variables.

Q3:

Going by the experience of $Q1$; the best guess, if we know the X value should be conditional expectation. Instead of depending on experience, we can derive this just like the earlier ones; but let us use experience. Thus let φ be the function

$$\varphi(a)=E(Y|X=a); \quad \text{if} \quad P(X=a)\neq 0;$$

$$\varphi(a)=0 \quad \text{if} \quad P(X=a)=0$$

We believe that $\varphi(X)$ (you make your own notation whether you want to use $\varphi(X)$ or composition notation φoX) should be the best guess.

Want to show that if you take any other square integrable function of X, say $\Psi(X)$; then

$$E\{[Y-\Psi(X)]^2\}\geq E\{[Y-\varphi(X)]^2\}.$$

Incidentally, you need not assume that $\Psi(X)$ is square integrable; you can dispose the above inequality in the other case easily. But do not bother about such minor details.

We proceed as in Q1. add and subtract $\varphi(X)$ in right side

$$E\{[Y - \Psi(X)]^2\} = E\{[Y - \varphi(X)]^2\} +$$

$$E\{[\varphi(X) - \Psi(X)]^2\} + 2E\{[Y - \varphi(X)][\varphi(X) - \Psi(X)]\} \quad (\dagger)$$

We need a very important fact now whose proof we shall see later. If you have any function of X, say $h(X)$ then

$$E[h(X)Y|X] = h(X)E(Y|X) \qquad (\clubsuit)$$

Note that in particular, taking $Y = 1$, we see $E[h(X)|X] = h(X)$.

{ This stands to reason. Easy to see $E[5Y|X] = 5E[Y|X]$.
Thus when you are told $X = a$, we see that
$h(X) = h(a)$ is just a number, so
$E[h(X)Y|X = a] = E[h(a)Y|X = a] = h(a)E[Y|X = a]$;
this being true for any number a, we conclude
$E[h(X)Y|X] = h(X)E[Y|X]$.
Unfortunately there are subtle issues to be settled. }

Returning to ($\dagger$); if $Z = [Y - \varphi(X)][\varphi(X) - \Psi(X)]$ then last term of dagger (apart from factor 2) equals $E(Z)$ which we understand now. In view of ($\clubsuit$), taking $h = \varphi - \Psi$ we see

$$E[Z|X] = [\varphi(X) - \Psi(X)] \;\; E\{[Y - \varphi(X)]|X\}$$

$$= [\varphi(X) - \Psi(X)] \;\; \{E[Y|X] - E[\varphi(X)|X]\}$$

If you look at the second term on right side you realize, by an application of dagger and definition of φ;

$$E[Y|X] - E[\varphi(X)|X] = \varphi(X) - \varphi(X) = 0$$

Thus $E(Z|X) = 0$. and hence

$$E(Z) = E\{E[Z|X]\} = E\{0\} = 0$$

Thus third term on RHS of ($\dagger$) is zero. Second term is nonnegative. This proves the required inequality. further equality holds in ($\dagger$) when and only when the following holds,

$$E\{[\varphi(X) - \Psi(X)]^2\} = 0.$$

Thus the best guess and expected penalty for this guess are

$$E(Y|X); \qquad E\{[Y - E(Y|X)]^2\} \quad (\bullet\bullet\bullet)$$

Calculus analogy:

Do you see that the first two questions are exactly like in Calculus? Of course, you need to interpret 'best' as your calculus teacher did.

Given a continuous function g on R and a point a what is the 'best constant function guess' for g near a?
Answer: the constant function $h(x) \equiv g(a)$.

Given a differentiable function g on R and a point a what is the 'best linear function guess' for g near a?
Answer: the tangent function at a; $h(x) = g(a) + g'(a)(x - a)$.

You can ask best quadratic guess etc. You come to Taylor expansion. In case of random variables too, we could have asked for the best quadratic guess $a + bX + cX^2$ for Y. We did not do.

In our case there is the third question $Q3$, you do not have this in calculus.

There is much more for conditional expectations. Those and (♣) will probably be discussed later. I want to show you some exciting things you can do with conditional probabilities which are extremely useful.

Chandrasekhar Model:

I have a huge supply of balls, as many as I want. I have two numbers: $0 < p < 1$ and $\lambda > 0$. Here is a game i play.

I start with a box having a certain number of balls.
Every morning, I take the balls in the box and for each ball decide to keep it or throw out. Having done that I add a certain number of balls.

Question: what happens in the long run?

of course, to make sense of the question and answer, you should know the mechanism of removing and adding balls. Here is how I remove:

Take a ball, toss coin; Heads up decide to remove, Tails up decide keep. Do this for each of the balls in the box.

Here is the mechanism of adding the balls:

Add $P(\lambda)$ many balls. This means, select an integer in such a way that chance of selecting n is

$$e^{-\lambda}\frac{\lambda^n}{n!}; \quad n = 0, 1, 2, \cdots .$$

Having selected an integer as above add so many balls. remember there is no limit for the number of balls being added but it is some finite number. So there never are infinitely many balls in the box.

Now the mechanism is completely specified. The only thing that needs to be told is: how did the game start? With how many balls did I start on day zero?

For simplicity let us assume that we started with zero balls.

later you will see that it does not matter. Even if you decide to roll a die and start with as many balls as the face that shows up, you will have exactly the same answer as we get below.

so let $p_k(n)$ be the probability of having k balls on day n. Thus if X_n is the number of balls on day n then you agree that it is a random variable.

$$p_k(n) = P(X_n = k).$$

Thus $p_k(0)$ is one if $k = 0$ and zero if $k \neq 0$.

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

because there is nothing to remove on day one, all the balls are those added that morning.

$$p_k(2) = P(X_2 = k) = \sum_{i=0}^{\infty} P(X_2 = k, X_1 = i)$$

$$= \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} P(X_2 = k|X_1 = i)$$

Let us denote

$$p_{ij} = P(X_2 = j | X_1 = i)$$

To have j balls tomorrow you should keep certain number of balls; this number can not exceed i (what you have) and also can not exceed j (what you want to have) and then add some to make total j. Thus

$$p_{ij} = \sum_{l=0}^{i \wedge j} \binom{i}{l} q^l p^{i-l} e^{-\lambda} \lambda^{j-l} \frac{1}{(j-l)!}.$$

In passing let us note that p_{ij} is also $P(X_{n+1} = j | X_n = i)$ whatever be n.

Returning to earlier calculation

$$p_k(2) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} \sum_{l=0}^{i \wedge k} \binom{i}{l} q^l p^{i-l} e^{-\lambda} \lambda^{k-l} \frac{1}{(k-l)!}.$$

$$= e^{-\lambda}\ e^{-\lambda}\ \lambda^k\ \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} q^l \sum_{i=l}^{\infty} p^{i-l} \lambda^{i-l} \frac{1}{(i-l)!}$$

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

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

Exactly the same argument shows

$$p_k(3) = e^{-\lambda(1+q+q^2)}\ [\lambda(1+q+q^2)]^k / k!.$$

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

Lo and behold

$$p_k(n) \to e^{-\lambda/p}\ (\lambda/p)^k\ \frac{1}{k!}.$$

Thus in the long run you expect to have $P(\lambda/p)$ many balls in the box. The In the long run, as usual means: rigorously 'as $n \to \infty$' and in practice 'for all large n'.

This model is called Chandrasekhar model. What is this model for and why are we removing and adding balls?

Well, actually balls are not balls and box is not box and day is not day and we are doing nothing either. Wow!

Imagine a Huge glass vessel filled with liquid. You put some coloured pollen particles (simply referred to as particles) into it. You fix, a small piece of volume in the middle of the vessel. This small piece of volume is our box (not the big thing). Day is not a day, but (1/200)th of a second. During every time interval, some of the particles (in the small volume, I fixed my attention on) leave that volume. Of course, some particles, from the Jar outside this volume, enter during this period.

What we discussed is precisely modelling this phenomenon. This model is due to the Indian Astrophysicist S Chandrasekhar. Data was actually collected, he was testing Brownian motion calculations in the theory of molecular fluctuations.

Why Poisson for the number of particles entering the volume under focus? There are so many particles in the Huge jar, each having a small chance of entering our region. Now you know what should be a good model for the number of particles entering.

The above process is an example of, what is called, Markov chain. This is probabilistic analogue of the deterministic 'dynamics' or 'motion' of physics. To describe motion of a system you need three things:

(0) where the motion is taking place
(1) how did the motion start
(2) what is the law governing the motion.

In our case the set where the motion is taking place is

$$S = \{0, 1, 2, 3 \cdots\cdots\}$$

This is called *state space.*

The starting state is zero. If we want we could have taken any other thing as mentioned earlier, even random. this is called *initial distribution.* In our case the initial distribution is the probability μ which puts mass one at the state zero. We could have taken any probability μ on S.

The law governing motion is the rule prescribing: if we are at i today, what are the chances of being at j tomorrow. in our case

$$p_{ij} = \sum_{l=0}^{i \wedge j} \binom{i}{l} q^l p^{i-l} e^{-\lambda} \lambda^{j-l} \frac{1}{(j-l)!}.$$

This matrix

$$((p_{ij})) \quad (i,j) \in S \times S$$

is called *transition matrix* for the motion.

Given these three ingredients, let X_n denote the state on day n. The sequence of random variables $(X_n : n \geq 0)$, is called *Markov Chain (MC)*. More precisely MC with state space S; initial distribution μ and transition matrix $((p_{ij}))$.

We are fortunate, we could make all the calculations needed. Not always is one so lucky as you will see later.

Questions like 'what happens in the long run' is a recurring theme in probability theory. After all, when we started developing probabilistic models of phenomena, we accepted that we could go wrong sometimes and what matters is 'how are we doing in general'.

Hardy Weinberg:

We shall consider another urn problem. Have two boxes: Mbox 1 and Fbox 1.

problem 1:

There are balls in each box and on each ball one of the following is written: AA, Aa, aa. Suppose these are in the proportion

$$AA : Aa : aa = u : v : w; \qquad u, v, w > 0; u + v + w = 1$$

in each box.

I Take one ball from each box and pick one letter from each ball.

What are the chances that I get AA: what are the chances I get Aa? what are the chances I get aa?

Here is a way of doing it. There are nine possible pairs of balls you could have picked up. You see

$$P(AA|\langle AA, AA\rangle) = 1; \qquad P(AA|\langle AA, Aa\rangle) = 1/2$$

$$P(AA|\langle Aa, Aa\rangle) = 1/4 \qquad P(AA|\langle AA, aa\rangle) = 0$$

Thus conditioning on the nine cases and doing the needed algebra, you get the following

$$AA : Aa : aa = p^2 : 2pq : q^2; \qquad p = u + \frac{1}{2}v; \;\; q = 1 - u = \frac{1}{2}v + w.$$

You can do in another way too, as one of you already suggested. Pick ball from Mbox and a letter from the ball then you get $A : a = p : q$ where $p.q$ are as above. Same happens if you pick from Fbox; and for the pair you get the above proportions.

Suppose now you start with Mbox 2 and Fbox 2 with these new and do the same. Then what happens? Nothing new, the same proportions are maintained, simply because $p^2 + pq = p$.

Thus from now on these proportions $p^2 : 2pq : q^2$ are maintained

problem 2:

Consider the same problem but initial proportions are as follows:

Mbox 1, $AA : Aa : aa = u : v : w$.
Fbox 1, $AA : Aa : aa = u^* : v^* : w^*$.

Simple calculations show that in both Mbox 2 as well as Fbox 2 the proportions are same and they are

$$AA : Aa : aa = pp^* : (pq^* + p^*q) : qq^*; \quad p = u + \frac{1}{2}v; \; p^* = u^* + \frac{1}{2}v^*$$

In Mbox 3 and Fbox 3 the proportions are

$$AA : Aa : aa = \alpha^2 : 2\alpha(1-\alpha) : (1-\alpha)^2$$

where

$$\alpha = pp^* + \frac{1}{2}(pq^* + p^*q) = \frac{1}{2}(p + p^*)$$

From now on these proportions $\alpha^2 : 2\alpha(1-\alpha) : (1-\alpha)^2$ are maintained.

Problem 3:

Consider similar problem as above: Mbox 1 and Fbox 1. But now only one letter A or a is written on balls in Mbox where as two letters AA or Aa or aa are written on balls of Fbox. Here are the proportions:

Mbox 1, $(A : a = p_0 : q_0)$
Fbox 1, $(AA : Aa : aa = u : 2v : w)$

As usual $0 < p_0 < 1; q_0 = 1 - p_0$ and $u, v, w > 0; u + 2v + w = 1$.

I pick one ball from Fbox 1 and pick a letter. What are the chances the letter is A and what are the chances it is a?

$$P(A) = p_1 \quad P(a) = q_1; \quad p_1 = u + v$$

These are the proportions in Mbox 2.

i pick one ball from Mbox 1 and one ball from Fbox 1. Pick one letter from each ball. What are the chances of AA? of Aa? of aa? Simple calculation shows the following.

$$P(AA) = p_0 p_1; \quad P(Aa) = p_0 q_1 + p_1 q_0; \quad P(aa) = q_0 q_1$$

These are the proportions in Fbox 2.

Start now with Mbox 2 and Fbox 2 and make Mbox 3 and Fbox 3 the same way.

Will the above proportions remain same?
No
Continue this procedure. What happens in the long run?
I leave it for you to argue that

$$p_n \to \alpha = \frac{1}{3} p_0 + \frac{2}{3} p_1$$

and

the limiting proportions are $A : a = \alpha : (1 - \alpha)$ *in Mbox and* $AA : Aa : aa = \alpha^2 : 2\alpha(1 - \alpha) : (1 - \alpha)^2$ *in Fbox.*

These three conclusions above are called **Hardy-Weinberg Laws** in genetics. What are the boxes, what are the letters and what is all this?

Humans have 23 pairs of chromosomes. They are nicely identified and numbered.

Let us consider the first 22 chromosome pairs. Each pair consists of similar things in structure but possibly different in chemical content.

Let us concentrate on one of these pairs and a particular location on this pair.

Let us assume that the chemical at this place on this chromosome can occur in two types A and a.

Thus the possible combination is one of the three types AA, Aa, aa. The first two problems refer to the proportions of these types and the phenomenon of how they change over generations. According to Mendelian laws child gets one letter from M and one from F. This is precisely what was happening in those problems.

The main conclusion is that the same proportions are maintained from the third generation, thus stability is attained soon in the population and you also know how the stable frequencies look like.

Of course, several assumptions are to be made, and one should also discuss the situation when some of the assumptions are not met. But this is not a course in genetics. Then why did I do this? Probability pervades all subjects. This is an important law and you should know. As some one said 'Hardy would be remembered more for this law than for his mathematical contributions'. Take this in spirit, not literally.

Now consider the 23rd pair of chromosomes. They are different both in structure and content. they are named X and Y. Those with XY are called M and those with XX are called F. Thus offspring always gets X from mother. The offspring is M if it gets Y from father; it is F if it gets X from father. The third problem refers to this.

We are considering a chemical present at a specific location on the X chromosome.

This chemical can occur in two types. Thus M has only one dose and F has two doses.

The H-W law gives us limiting proportions. Here is a specific example of such chemical: one that controls 'colour blindness' (another is Hemophilia). Think of a as the chemical causing colourblindness and A is the chemical for normal vision. Suppose a proportion 1 in 1000 of males are colour bind. What about females? Luckily the chemical A is dominant, its presence makes a in effective. Thus F would be colour blind only when aa. Thus among F only 1 in 1000000 will be colour blind. We are assuming limiting proportions, but do not worry, so many generations have elapsed; you can safely do so.

We shall not discuss any more genetics. We shall see other areas.

Rnadom Walk:

We shall discuss one more example of Markov chain described earlier. Now the state space is the set of integers Z. The law of motion is the following. if you are at an integer, toss fair coin, Heads up move one step forward, Tails up move one step backward.

Thus transition probabilities are

$$p_{i,i-1} \quad = \quad p_{i,i+1} \quad = \quad \frac{1}{2}; \qquad p_{ij} = 0 \quad j \neq i \pm 1.$$

On day zero we start at 0.

Question: is there a chance of returning to zero later?

Answer: Yes, at least 1/2 because HT and TH will lead you back there.

Question: is there a chance of NOT returning? Alternatively, can we be sure of returning?

This is the question we ask. Of course, you might wonder; why did you not ask 'what happens in the long run'. Well, nothing happens; there will be lots and lots fluctuations. You will realize later.

Let us use notation and make the question precise.
[There is a point on which I shall be silent, simply because you would probably not realize it.]
Shall discuss later. Let $X_0 = 0$ and X_n be our state on day n for $n \geq 1$. The question thus is

$$P(X_n = 0 \quad \text{for some} \quad n \geq 1) = 1?$$

Let us denote

$$p^{(n)} = P(X_n = 0), \quad n \geq 1; \quad p^{(0)} = 1$$

We can easily evaluate this.

$$p^{(2n+1)} = 0; \qquad\qquad p^{(2n)} = \binom{2n}{n} \frac{1}{2^{2n}}$$

Unfortunately these events $(X_n = 0)$ are not disjoint and so will not lead to an immediate answer to our problem. let us disjointify our events.

$$S_1 = (X_1 = 0); \qquad S_2 = (X_1 \neq 0; X_2 = 0)$$

In general

$$S_n = (X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} \neq 0;\ X_n = 0)$$

These events are disjoint and it is clear that

$$(X_n = 0 \quad \text{for some} \quad n \geq 1) \quad = \quad \bigcup_1^\infty S_n$$

You might be wondering why I am bothering to write S_{2n+1}, is it not empty? Yes, but I do not want to keep track of such details at this stage and do calculations that might be useful more generally. As a consequence, if we define

$$f^{(n)} = P(S_n), \quad n \geq 1; \qquad f^{(0)} = 0$$

then our problem can be reformulated as

$$\sum f^{(n)} \; = \; 1?$$

Unfortunately these $f^{(n)}$ are usually difficult to calculate, though in this example we can use reflection principle and calculate. That leads to a powerful method, but with limited applications. So we follow a different route.

We show the following:

■ Theorem: $\sum f^{(n)} = 1$ if and only if $\sum p^{(n)} = \infty$. (♠) ■

Accept this for now. Let us see how this helps us. This reduces our problem to decide whether the following series converges or not.

$$\sum \binom{2n}{n} \frac{1}{2^{2n}}$$

But how do we understand these numbers? This is where we take the help of James Stirling. Stirling's formula says

Theorem :

$$n! \; \sim \; \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}.$$

What is the notation? Suppose that (a_n) and (b_n) are sequences of strictly positive numbers. We use

$$a_n \sim b_n \quad \Longleftrightarrow \quad \frac{a_n}{b_n} \to 1.$$

From the definition you can show the following.

$$a_n \sim b_n; c_n \sim d_n \quad \Rightarrow \quad a_n c_n \sim b_n d_n; \ \frac{a_n}{c_n} \sim \frac{b_n}{d_n}$$

As a result

$$\frac{2n!}{n!n!}\,\frac{1}{2^{2n}} \quad \sim \quad \frac{1}{\sqrt{\pi n}}$$

It is also easy to see that

$$a_n \sim b_n \quad \Longrightarrow \quad \left(\sum a_n < \infty \leftrightarrow \sum b_n < \infty\right)$$

This can be seen by just noting that after some stage

$$\frac{1}{2} \ < \ \frac{a_n}{b_n} \ < \ 3$$

Now recall

$$\sum \frac{1}{\sqrt{n}} = \infty.$$

As a result we conclude that

$$\sum p^{(n)} = \infty$$

This solves our problem. Thus

We are sure to return to zero.

Two dimensions:

The method is so powerful we can answer in higher dimensions too. Consider the state space

$$S = \{(i,j) : i,j \in Z\} \quad \subset \quad R^2$$

From any point choose one coordinate axes and on that move forward or backward one unit. More precisely from (i,j) move to one of the four points $(i, j \pm 1)$ and $(i \pm 1, j)$ each with probability $1/4$.

We start at $(0,0)$.

Question: Are you sure to return to $(0,0)$?

Let us interpret the zero of the above calculations as the zero of R^2 then (♠) remains true. In other words let us now make the analogous definitions.

$$p^{(n)} = P\{X_n = (0,0)\}, \quad n \geq 1; \quad p^{(0)} = 1$$

$$S_1 = \{X_1 = (0,0)\}; \qquad S_2 = \{X_1 \neq (0,0); X_2 = (0,0)\}$$

In general

$$S_n = \{X_1 \neq (0,0), X_2 \neq (0,0), \cdots, X_{n-1} \neq (0,0);\ X_n = (0,0)\}$$

$$f^{(n)} = P(S_n), \quad n \geq 1; \qquad f^{(0)} = 0$$

With this notation (♠) still holds. Thus to answer our question we need to find our whether $\sum p^{(n)}$ converges or not.

As earlier, we can argue to see

$$p^{(2n+1)} = 0$$

$$p^{(2n)} = \sum_{k=0}^{n} \frac{(2n)!}{k!k!(n-k)!(n-k)!} \ \frac{1}{4^{2n}}.$$

This is because to be at origin on day $2n$, we should have made a certain number (may be zero, may be n, but not more) of right moves and same number of left moves and in the remaining $(2n-2k)$ days we should make $(n-k)$ top moves and same number of down moves. Thus

$$p^{(2n)} = \binom{2n}{n} \sum_{k=0}^{n} \frac{n!n!}{k!k!(n-k)!(n-k)!} \ \frac{1}{4^{2n}}.$$

$$= \binom{2n}{n}\binom{2n}{n} \frac{1}{4^{2n}} \ \sim \ \frac{1}{\pi n}.$$

Thus again the sum $\sum p^{(n)} = \infty$. Thus

We are sure to return to the origin.

Three dimensions:

Now consider the state space

$$S = \{(i,j,k) : i,j,k \in Z\} \quad \subset \quad R^3$$

Law of motion is as follows. Choose one coordinate axes and in that direction move one step forward or one step backward. In other words, from (i,j,k) you move with equal probability to one of the six points:

$$(i \pm 1, j, k), \quad (i, j \pm 1, k), \quad (i, j, k \pm 1).$$

We start at $(0,0,0)$.

You are mature enough now to make the earlier notation meaningful by interpreting zero as origin $(0,0,0)$.

Let us pretend that (♠) is so powerful, it is valid even now.

To be at origin on day $2n$ you need to make certain number k of moves in X^+ direction and same number in X^- direction; a certain number l in Y^+ direction and same number in Y^- direction; make sure $k + l \le n$ and in the other $2n - 2k - 2l$ days half Z^+ and half Z^- moves. Thus

$$p^{(2n)} = \sum_{0 \le k,l,k+l \le n} \frac{(2n)!}{k!\; k!\; l!\; l!\; (n-k-l)!\; (n-k-l)!} \; \frac{1}{6^{2n}}$$

$$= \binom{2n}{n} \frac{1}{2^{2n}} \frac{1}{3^n} \sum_{0 \le k,l,k+l \le n} \left[\frac{n!}{k!\; l!\; (n-k-l)!} \right]^2 \frac{1}{3^n}$$

If you have positive numbers a_i and $\max a_i \le M$ then

$$\sum a_i^2 \le M \sum a_i$$

Thus if

$$\max \left\{ \frac{n!}{k!\; l!\; (n-k-l)!} : 0 \le k,l,k+l \le n \right\} = M_n$$

then

$$\sum_{0 \le k,l,k+l \le n} \left[\frac{n!}{k!\; l!\; (n-k-l)!} \right]^2 \frac{1}{3^n} \le M_n \sum_{0 \le k,l,k+l \le n} \left[\frac{n!}{k!\; l!\; (n-k-l)!} \right] \frac{1}{3^n}$$

But the last sum is just trinomial expansion of

$$\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right)^n = 1$$

Thus

$$p^{(2n)} \leq \binom{2n}{n} \frac{1}{2^{2n}} \ \frac{1}{3^n} M_n$$

We know

$$\binom{2n}{n} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}$$

Remembering that the multinomial terms are the largest at

$$k = l = (n - k - l) = n/3$$

(we are careless here) we see

$$\frac{1}{3^n} M_n \leq \frac{1}{3^n} \frac{n!}{(\frac{n}{3})! \ (\frac{n}{3})! \ (\frac{n}{3})!}$$

$$\sim \ \frac{1}{3^n} \ [\sqrt{2\pi} e^{-n} n^{n+(1/2)}] \ [\sqrt{2\pi} e^{-n/3} (n/3)^{(n/3)+(1/2)}]^{-3}$$

$$= \frac{1}{2\pi} \ 3^{3/2} \ \frac{1}{n}$$

where we used Stirling. Thus ultimately

$$p^{(2n)} \leq \alpha_n; \qquad \qquad \alpha_n \sim C n^{-3/2}$$

for some constant C. Since $\sum n^{-3/2}$ and hence $\sum \alpha_n$ and hence $\sum p^{(2n)}$ converges we conclude the following.

In three dimensions there is a positive chance of NOT returning back to the origin.

We were careless at one point: $n/3$ may not be integer and hence saying that $(n/3, \ n/3, \ n/3)$ term is largest does not make sense. How do you rectify it?

Well, the argument above definitely shows that sum of all $p^{(2n)}$ where n is multiple of 3, is finite.

Consider the sum of $p^{(2n)}$ over n of the form 1 mod(3). That is, over all n of the form $3m + 1$. Now the max term is attained at $(m, m, m + 1)$ Do similar Stirling analysis and show sum is finite.

Then consider n of the form 2 mod(3). The multinomial term for this case attains maximum at $(m, m+1, m+1)$ and again show this sum is finite.

Thus you deduce that $\sum p^{(2n)}$ is finite. AAha, done!