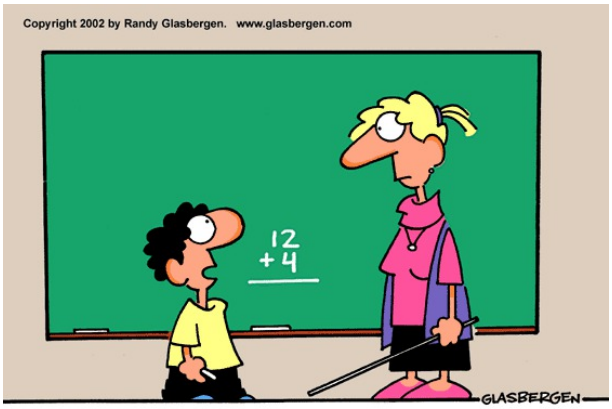

"Do I get partial credit for simply having the courage to get out of bed and face the world again today?"

CMI/BVR Probability notes 4

Poisson distribution:

Before discussing joint distribution, here is an interesting question of practical importance: Suppose I toss a coin very very large number of times, but the chance of heads is very very small. How do we describe the number of heads, that is, its distribution?

Of course, you can say it is binomial. But what is the worth of this answer? Suppose the chance of heads is 0.000001 and you toss the coin 1000000 times. Do you still say that the chances of 3 heads is

$$\binom{1000000}{3}(0.000001)^3(1-0.000001)^{999997}$$

Is it not better to approximate it by an easily understandable number?

Here is the thought process. Let us say p is very small and n is very large. Then the chance of zero heads is

$$(1-p)^n = \left(1-\frac{\lambda}{n}\right)^n \sim e^{-\lambda}; \quad \lambda = np.$$

Thus if np is a reasonable number then we can arrive at an approximation.

■ **Theorem:** Suppose $X_n \sim B(n, p_n)$. Assume $np_n \to \lambda$; $0 < \lambda < \infty$. Then,

$$P(X_n = k) \to e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, 2, 3, \cdots .$$

Proof is simple. Fix k. We need to calculate limit of

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} =$$

$$= \frac{[np_n][(n-1)p_n][(n-2)p_n]\cdots[(n-k+1)p_n]}{k!}\ \left(1 - \frac{np_n}{n}\right)^{n-k} \quad (\bullet)$$

Remember that k is fixed so that there are fixed number of terms in the product above. if you denote $\lambda_n = np_n$ then $\lambda_n \to \lambda$ so that

$$\left(1 - \frac{np_n}{n}\right)^{n-k} = \left(1 - \frac{\lambda_n}{n}\right)^{n}\ (1-p_n)^{-k}$$

$$\longrightarrow e^{-\lambda} \cdot 1. \qquad (\spadesuit)$$

Also for each i

$$(n-i)p_n \ = \ np_n - ip_n \ \to \ \lambda - 0 \ = \ \lambda. \quad (\clubsuit)$$

Using ($\spadesuit$) and ($\clubsuit$) in ($\bullet$) we get the stated result. ■

Note that these limiting numbers add to one. Thus we can think of a random variable X taking values and probabilities as follows.

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

Such a random variable is called **Poisson** random variable and the probabilities are called **Poisson probabilities.** More precisely, one says X is Poisson with parameter λ, denoted $X \sim P(\lambda)$.

You should be under no illusion (at this stage of our life) that the random variables X_n of the theorem are converging, in some sense, to a random variable X. We are only saying that the probabilities for each value k converge to a limiting value and so we thought of an ideal distribution and named it Poisson distribution. A random variable with such a distribution is Poisson variable.

Thus the practical implication is the following. If p is very small and n is very large then we can approximate the binomial probabilities with Poisson

probabilities described above, where λ is the mean of your binomial variable, namely $\lambda = np$.

The theorem above is precise. The interpretation of the above para is not precise simply because we used terms like very large and very small. You very well know there are no quantifications for these terms. This is a practical advise: To calculate $P(X_n = k)$ you can as well calculate $P(X = k)$ where $X \sim P(\lambda)$ where $\lambda = np$.

Suppose you have sequence of numbers (a_n). After all, when we see that 'a_n is getting pretty close to $1/2$ as n becomes large' we agreed to say $a_n \to 1/2$.

Are there error bounds and what is the worth of this approximation? Yes, there are error bounds but we shall not discuss. Yes, it is worth. This distribution seems to fit some observed phenomena.

Consider disintegration of a radio active material. How many particles disintegrate in one hour? You can put a Geiger counter and count the number. it is found that the data follows this $P(\lambda)$ probabilities with very good accuracy. The reason is simple. Each particle has a very very very small chance of jumping out because the other particles are pulling it. But there are very very very large number of particles in the material. So sooner or later one particle does succeed in freeing itself. the number particles that free themselves in an hour is a Poisson variable. of course the number λ depends on the material, its half life.

Consider number of road accidents in a month at a busy intersection. A Poisson variable is found to be best fit for observed data. After all, a very very very large number of cars do pass through this intersection. But the chances that a car is involved in an accident is very very very small. So the number of successes (accident being considered as success!) is a Poisson variable.

There are many other data sets that follow these Poisson probabilites.

joint distribution:

Let (Ω, p) be a probability space. If you have two random variables X and Y then by looking at their distributions separately, does not give us any idea of possible connections between these variables.

For example, If X is the number of heads and Y the number of Tails in two tosses of a fair coin then both are $B(2, 1/2)$ variables. It is not clear that $X + Y \equiv 2$ when you look at their distributions separately, as observed last time. Here is how to look at them jointly.

Method 1:

Just like one r.v., suppose we plot the values of the pair (X, Y) along with their corresponding probabilities.

values	probabilities
$(0, 2)$	$1/4$
$(1, 1)$	$1/2$
$(2, 0)$	$1/4$

Then it is clear that numbers in each pair add to two.

Method 2:

You can present the same table as a bivariate table as follows. the values of X are in the top horizontal margin. The values of Y are in the left vertical margin. The (i, j)-th entry in the table is the probability that $(X = i, Y = j)$. The bottom horizontal margin are the column totals and the right vertical margin are the row sums.

If you read the top and bottom margins then you see the distribution of X. If you read the left and right vertical margins then you see the distribution of Y.

$Y \backslash X$	0	1	2	total
0	0	0	$1/4$	$1/4$
1	0	$1/2$	0	$1/2$
2	$1/4$	0	0	$1/4$
total	$1/4$	$1/2$	$1/4$	1

Even now it is clear that only those pairs of values that add to two have positive probability and others have probability zero.

Method 1 has the advantage that only those pairs which have non-zero probability are listed, unnecessary pairs are not listed. However if some one

asks you what are the possible values of X alone; you need to see each pair and pick up the first coordinate, time consuming. Of course, in the above example it is simple but in general it is not so.

Method 2 appears a neater presentation, though there are several zeros in the matrix of probabilities. However you can immediately understand, without any calculation, the possible values of X and their probabilities: just read the top and bottom margins. Similarly for Y. Thus by looking at this table, you can not only understand the pair (X, Y) but also X alone and Y alone too without any further work. You will be able to detect some interesting phenomena too by just staring at the table. We see now.

Consider tossing a fair coin four times. let X denote the number of heads in the first two tosses and Y the number of heads in the last two tosses.

We shall not describe method 1; it has nine tuples with corresponding probabilities. here is method 2.

$Y\backslash X$	0	1	2	total
0	1/16	1/8	1/16	1/4
1	1/8	1/4	1/8	1/2
2	1/16	1/8	1/16	1/4
total	1/4	1/2	1/4	1

The table immediately tells you another interesting feature: product of the marginals gives the corresponding entry in the matrix. More precisely

$$i\text{-th row } j\text{-th column entry } = (i\text{-th row total}) \cdot (j\text{-th column total}).$$

Such a feature would not be easily detectable if we used method 1.

Suppose that (Ω, p) is a probability space. Suppose $X_1, X_2, \cdots, X_k$ are k random variables defined on this space. Their **joint distribution** is a table giving all possible values $(a_1, a_2, \cdots, a_k)$ of this k-tuple of variables and against each value its probability, that is,

$$P(\omega : X_1(\omega) = a_1; \cdots, X_k(\omega) = a_k).$$

table: $(a_1, \cdots, a_k) \qquad P(a_1, \cdots, a_k)$ $\qquad (\star)$

However, if there are two variables, we prefer the bivariate table shown above, from which several features can be immediately extracted by inspection.

An analogue of the bivariate table for the k-dimensional distribution is the following. First collect the set D_1 of possible values of X_1; the set D_2 of possible values of X_2; and so on. Thus let D_i be the set of possible values of X_i for $1 \leq i \leq k$. Then the k-variate table is

$$P(a_1, \cdots, a_k): \quad (a_1, \cdots, a_k) \in D_1 \times \cdots \times D_k \quad (\star\star)$$

Given joint distribution as above the marginal distribution is the marginal totals. For example the first marginal total gives

$$P(X_1 = \alpha) = \sum_{a_2, \cdots, a_k} P(\alpha, a_2, \cdots, a_k): \qquad \alpha \in D_1. \quad (\spadesuit)$$

Similarly you can make the i-th marginal.

Note that $(\spadesuit)$ indeed gives the distribution of the random variable X_1. This is because, the event $(X_1 = \alpha)$ is the disjoint union of the events

$$\{(X_1 = \alpha, X_2 = a_2, \cdots, X_k = a_k): \quad a_2 \in D_2, \cdots, a_k \in D_k\}$$

Thus, this marginal is called the marginal distribution of X_1. of course, marginal distribution of X_1 is indeed the distribution of X_1. The adjective 'marginal' draws your attention to the fact that there are other variables and they had a joint distribution and by looking at the 'appropriate' marginal we got the distribution of X_1.

Digression: table

We have been using the word table to describe distributions. It is not necessary to use this word.

Distribution of a random variable X is the function defined on R as follows:

$$f(a) = P\{\omega : X(\omega) = a\} \qquad a \in R$$

This is called *probability mass function* (in order not to be confused with another function we come across later) of X. Since Ω is countable we conclude that X takes only countably many values and the event in the display above is empty set for all except countably many numbers a. Our earlier

description used only these values a in the table.

Similarly, if we have k random variables $X_1, \cdots, X_k$ then their distribution is the function defined on R^k as follows:

$$f(a_1, \cdots, a_k) = P\{\omega : X_i(\omega) = a_i \ \forall \ i\} \qquad (a_1, \cdots, a_k) \in R^k$$

This is called *joint probability mass function* (in order not to be confused with another function we come across later) of the k variables $X_1, \cdots, X_k$. Since Ω is countable, we see that for only countably many k-tuples, the event in the display above is non-empty. Our earlier description used only these tuples in the table. In the second method we used a 'box' of values for the tuple. Think about it.

expectation of $\varphi(X_1, \cdots, X_k)$:

Suppose we have r.v. $X_1, \cdots, X_k$ defined on a probability space (Ω, p). Suppose that $\varphi : R^k \to R$ is a function. We can define a random variable Z on Ω as follows.

$$Z(\omega) = \varphi(X_1(\omega), \cdots, X_k\ (\omega)) = \varphi o(X_1, \cdots, X_k)(\omega).$$

We can think of calculating $E(Z)$ in several ways.

In what follows we assume that all sums, we are dealing with, are absolutely convergent.

■ (**A**) We can calculate the distribution of Z; say takes values (z_n) with respective probabilities (α_n). Then compute

$$E(Z) = \sum_n z_n \alpha_n$$

(**B**) You can calculate the joint distribution of $(X_1, \cdots, X_k)$; say this tuple takes values $\{(a_1, \cdots, a_k)\}$ with respective probabilities $\{p(a_1, \cdots, a_k)\}$. then compute

$$E(Z) = \sum_{(a_1, \cdots, a_k)} \varphi(a_1, \cdots, a_k) p(a_1, \cdots, a_k)$$

(**C**). Do not calculate any distribution. Look at sample space and compute

$$E(Z) = \sum_\omega \varphi(X_1(\omega), \cdots, X_k(\omega)) p(\omega) = \sum_\omega Z(\omega) p(\omega)$$

Are these equivalent? Do they give same answer?

Note that (**A**) is definition and the others are not.

Yes. they all give the same answer. We had already seen, in one variable set-up that (**A**) and (**C**) are equivalent.

Similar proof shows that (**B**) and (**C**) are also equivalent. Here is how. Start with (**C**). Do the sum in two ways. First for each fixed tuple $(a_1, \cdots, a_k)$; sum or subtotal only over those ω in the set

$$\{X_1 = a_1, \cdots, X_k = a_k)$$

Then add all these subtotals, you get (**B**). ∎

if you know a little analysis, you can actually show that if one series above converges absolutely, then so do the others. Thus absolute convergence of any one of the above three series can be taken towards the existence of $E(Z)$ and then any one of the three series above can be used as definition of $E(Z)$.

The equivalence of the above three will be referred to as **substitution formula** or **change of variable formula**.

independence:

We say k random variables are **independent** if

$$P(X_i = a_i : 1 \leq i \leq k) = \prod_1^k P(X_i = a_i)$$

for each k-tuple of numbers. This amounts to saying that the entry at any place in the table of probabilities equals product of the corresponding marginal numbers.

Suppose that $X_1, \cdots, X_k$ are independent. Suppose B_i is any set of possible values of X_i. Then using the fact that probability of union is sum of probabilities for disjoint events;

$$P(X_i \in B_i \ \forall \ i) =$$

$$P\{(X_1, \cdots, X_k) = (b_1, \cdots, b_k) \text{ for some } (b_1, \cdots, b_k) \in \overset{k}{\underset{1}{\times}} B_i\}$$

$$= \sum_{(b_1,\cdots,b_k)\in \overset{k}{\underset{1}{\times}} B_i} P\{(X_1,\cdots,X_k)=(b_1,\cdots,b_k)\}$$

$$= \sum_{(b_1,\cdots,b_k)\in \overset{k}{\underset{1}{\times}} B_i} P(X_1=b_1)P(X_2=b_2)\cdots P(X_k=b_k)$$

Performing the addition over $b_1 \in B_1$ and then over $b_2 \in B_2$ etc successively, we see

$$P(X_i \in B_i \ \forall \ i) = P(X_1 \in B_1)P(X_2 \in B_2)\cdots P(X_k \in B_k).$$

The definition of independence corresponds to the special case of the above equation where each B_i is a singleton.

Thus we have the following:

■ The r.v. $X_1, X_2, \cdots, X_k$ are independent iff for any $B_1, B_2, \cdots, B_k \subset R$

$$P(X_i \in B_i \ \forall \ i) = \prod_1^k P(X_i \in B_i) \qquad ■$$

of course, earlier we used the phrase 'B_i is a set of possible values of X_i' and in the above statement we did not use this phrase. If B_1 is an arbitrary subset of R then $B_1 = C_1 \cup D_1$ where C_1 consists of all numbers which are possible values of X_1 and D_1 is the set of remaining numbers of B_1. Then $P(X_1 \in B_1) = P(X_1 \in C_1)$.You can use this fact to see that omission of the phrase causes no problem.

By taking some of the sets B_i to be R we see the following:

■ If $X_1, \cdots, X_k$ are independent then any sub collection is also independent. For example X_2, X_4, X_7 are independent. ■

■ If X and Y are independent random variables on a probability space, $E(X)$ and $E(Y)$ defined; Then $E(XY)$ is also defined and

$$E(XY) = E(X)E(Y)$$

Let us not spend time on existence pat. Towards the proof of the equality, use the function $\varphi(a,b) = a.b$ in substitution rule to see

$$E(XY) = \sum_{a,b} \varphi(a,b)P(X=a, Y=b)$$

Use independence

$$= \sum_{a,b} a\ b\ P(X=a)P(Y=b)$$

sum w.r.t. b;

$$= \sum_{a} aP(X=a)E(Y) \quad = \quad E(X)E(Y).$$

Covariance, correlation:

if two random variables X, Y are independent then we have

$$E(XY) = E(X)E(Y)$$

It appears reasonable to take $E(XY) - E(X)E(Y)$ as a measure of non-independence or relatedness.

We define **covariance** between X, Y by

$$Cov(X,Y) = \sigma_{X,Y} = E(XY) - E(X)E(Y)$$

$$\text{correlation}(X,Y) = \rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Thus if X, Y are independent then $cov(X,Y) = 0$. However $cov(X,Y) = 0$ does not mean that they are independent.

Example: Let X take values $\{-2, -1, 1, 2, \}$ each with probability $1/4$. Let $Y = X^2$. then they are not independent because

$$P(X=1, Y=1) = P(X=1) = \frac{1}{4}$$

$$P(X=1)P(Y=1) = \frac{1}{4}\frac{1}{2} = \frac{1}{8}$$

However

$$E(X) = 0; \quad E(XY) = E(X^3) = 0; \quad E(XY) - E(X)E(Y) = 0$$

We defined variance of X as $\sum (x_i - \theta)^2 p_i$. Here X takes values (x_i) with probabilities (p_i); and θ is $E(X)$. Note

■ $\quad Var(X) = E(X^2) - [E(X)]^2 = E\{[X - E(X)]^2\} \quad (\dagger) \quad$ ■

Here first equality is because

$$\sum (x_i - \theta)^2 p_i = \sum x_i^2 p_i - 2\theta \sum x_i p_i + \theta^2$$

$$= E(X^2) - 2\theta^2 + \theta^2 = E(X^2) - [E(X)]^2$$

We have used substitution formula to identify $\sum x_i^2 p_i$ with $E(X^2)$.

the second equality in (†) follows from linearity of expectation

$$E\{(X-\theta)^2\} = E\{X^2 - 2\theta X + \theta^2\}$$

$$= E(X^2) - 2\theta E(X) + \theta^2 = E(X^2) - \theta^2.$$

Suppose the pair (X, Y) takes values $\{(x_i, y_j)\}$ with probability $\{p_{i,j}\}$. suppose $E(X) = \theta$ and $E(Y) = \mu$. then according to definition

$$cov(X, Y) = E(XY) - \theta\mu$$

This is also same as

$$\blacksquare \quad Cov(X, Y) = E\{(X-\theta)(Y-\mu)\} = \sum_{i,j} x_i y_j p_{i,j} - \theta\mu \quad (*) \quad \blacksquare$$

The first equality above is again by linearity of expectation.

$$E\{(X-\theta)(Y-\mu)\} = E\{XY - \theta Y - \mu X + \theta\mu\}$$

$$= E(XY) - \theta E(Y) - \mu E(X) + \theta\mu = E(XY) - \theta\mu$$

For the last equality of $(*)$ we only need to identify $\sum x_i y_j p_{i,j}$ with $E(XY)$, use substitution formula.

Note also that

$$\blacksquare \quad \sigma_X^2 = \sigma_{X,X} = Cov(X, X) \quad \blacksquare$$

Here is an important formula which will be used soon.

Variance of sum:

$\blacksquare$ for random variables $X_1, X_2, \cdots, X_k$ we have

$$Var(\sum_1^k X_i) = \sum_1^k Var(X_i) + 2\sum_{i<j} cov(X_i, X_j).$$

If the random variables are independent, then variance of sum equals sum of variances. $\blacksquare$

Proof is simply high school algebra and linearity of expectation.

$$Var(\sum_{1}^{k} X_i) = E\{(\sum_{1}^{k} X_i)^2\} - \{\sum E(X_i)\}^2$$

$$= \sum_{1}^{k} E(X_i^2) + 2\sum_{i<j} E(X_i X_j) - \sum_{1}^{k}\{E(X_i)\}^2 - 2\sum_{i<j} E(X_i)E(X_j)$$

$$= \sum_{1}^{k} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

the last statement regarding independent variables is immediate because in that case the covariance terms equal zero.

Negative Binomial:

Consider tossing a coin till third head and X is number of failures.
We define some more random variables on the sample space as follows.

W_1 : Number of Tails before first Head.

W_2 : Number of Tails between first and second Head.

W_3 : Number of Tails between second and third Head.

These are called waiting times. Clearly

$$X = W_1 + W_2 + W_3$$

We shall argue that these are independent and each is $G(p)$. Intuitively it is clear. W_1 is number of tails before a Head and is hence $G(p)$. The tosses being independent; after the first head everything starts independently and we are now waiting again for a Head. Thus W_2 is again $G(p)$ and is independent of W_1. similarly for W_3.

Such a thought process is necessary to develop intuition. Unfortunately, by itself, achieves nothing mathematically. Intuition should be substantiated by mathematics.

You might even feel that the following argument is rigorous: W_1 is $G(p)$ because it is waiting time for Head to appear, how do future tosses matter for

this? Yes, right, but the last phrase 'how do future tosses matter' is not for *you* to decide; it is a matter that is to be justified. Obviously the following tw experiments are different: Experiment 1: toss a a coin till you get heads (and then count Tails). Experiment 2: toss till you get third head (and then count number of tails before first head).

The point is that you should recognize that mathematical justification is lacking. Once you realize this, matters are simple; maths is NOT difficult if you know what needs to be done. Most of the time math is difficult because we do not understand what we need to do, but we proceed aimless.

Let us see the joint distribution of these r.v. The possible values of this triple (W_1, W_2, W_3) consists of all triples of non-negative integers. For such a triple, denoting $(1-p) = q$, the event

$$P(W_1 = a, W_2 = b, W_3 = c) = q^a pq^b pq^c p$$

because this event consists of only one outcome $T^a HT^b HT^c H$.

Let us calculate marginals.

$$P(W_1 = a) = \sum_{b \geq 0, c \geq 0} q^a pq^b pq^c p = q^a p$$

where we have summed successvely w.r.t. b and then w.r..t c. Similarly

$$P(W_2 = b) = \sum_{a \geq 0, c \geq 0} q^a pq^b pq^c p = q^b p$$

$$P(W_3 = c) = \sum_{a \geq 0, b \geq 0} q^a pq^b pq^c p = q^c p$$

It is now clear

$$P(W_1 = a, W_2 = b, W_3 = c) = P(W_1 = a)P(W_2 = b)P(W_3 = c)$$

Thus both the statements, each $W_i \sim G(p)$ and $\{W_i\}$ are independent follow.

As a result we see

$$E(X) = 3\frac{q}{p}; \quad Var(X) = 3\frac{q}{p^2}$$

of course the first equation does not need independence. The second needs independence. We used the formula for mean and variance of $G(p)$ calculated

earlier. Thus no fresh calculations are needed to obtain mean and variance of X.

In general if $X \sim NB(r,p)$ then

$$E(X) = r\frac{q}{p}, \qquad Var(X) = r\frac{q}{p^2}$$

Multinomial distribution:

Roll a dice 30 times. Chance of face i appearing is p_i for $1 \leq i \leq 6$. Thus $p_i \geq 0$ and $\sum p_i = 1$.

Unless stated otherwise repetitions of an experiment are to be regarded as independent repetitions. Thus here we roll the die independently. remember this phrase comes into action when you assign probabilities to outcomes.

Let X_i be the number of times face i appears. We want to find the joint distribution of $(X_1, \cdots, X_6)$. Values of this are all possible tuples of non-negative integers $(n_1, \cdots, n_6)$ such that $\sum n_i = 30$.

You can see that for a given tuple like above, there are

$$\frac{30!}{n_1! \cdots n_6!}$$

outcomes and each outcome has probability

$$p_1^{n_1} p_2^{n_2} \cdots p_6^{n_6}.$$

Thuis distribution of $(X_1, \cdots, X_6)$ is

Values: $(n_1, \cdots, n_6)$: each $n_i \geq 0$ integer and $\sum n_i = 30$.

Probabilities: $P(n_1, \cdots, n_6) = \dfrac{30!}{n_1! \cdots n_6!} p_1^{n_1} p_2^{n_2} \cdots p_6^{n_6}$.

These are called **Multinomial probabilities** and the distribution is called **multinomial distribution**. The tuple $(X_1, \cdots, X_6)$ is called multinomial random vector.

Why are these called so? because these numbers are precisely terms in the expansion of

$$(p_1 + \cdots + p_6)^{30}$$

.

What is the distribution of X_1? Well, you can ignore the finer details of the basic experiment and simply pretend that it is two outcome experiment: say success if face 1 appears otherwise failure. thus the basic experiment has chance of success p_1 and failure $(1-p_1)$. We are repeating this experiment 30 times. Hence

$$X_i \sim B(30, p_1)$$

What is the marginal distribution of X_1? of course, you can say that marginal distribution is same as distribution and hence it is as above. It is perfectly correct to say so.

Let us calculate the marginal distribution of X_1? For each k, $1 \leq k \leq 30$ we should sum the above probabilities over $(n_2, \cdots, n_6)$.

$$P(X_1 = k) = \sum_{n_2, \cdots, n_6} \frac{30!}{k! \cdots n_6!} p_1^k p_2^{n_2} \cdots p_6^{n_6}.$$

Here the sum is over all tuples adding to $30-k$

$$= \frac{30}{k!(30-k)!} p_1^k \sum_{n_2, \cdots, n_6} \frac{(30-k)!}{n_2 \cdots n_6!} p_2^{n_2} \cdots p_6^{n_6}.$$

$$= \frac{30}{k!(30-k)!} p_1^k (1-p_1)^{30-k}$$

as expected.

More generally suppose you have a dice with r faces and chance of face i appearing in a throw is p_i. Roll the dice n times and denote X_i as the number of times face i appears. Then $(X_1, \cdots, X_r)$ has the following distribution.

Values: $(n_1, \cdots, n_r)$: each $n_i \geq 0$ integer and $\sum n_i = 30$.

Probabilities: $P(n_1, \cdots, n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$.

This is called multinomial with parameters $(n; p_1, \cdots, p_r)$.

Some people refer to this as Multinomial with parameters $(n; p_1, \cdots, p_{r-1})$. There is no confusion regarding which notation is being used. In this second notation these p_i do not add to one. So understand that there are r possibilities and the last possibility has probability

$$(1 - \sum_1^{r-1} p_i).$$

A coin is also a dice — it is a two faced dice.

Also, when you say dice, without any adjective regarding faces, it means usual dice with six faces.

Also, when you do not specify probabilities for the faces, it means that all faces are equally likely.

Also, when you say roll the dice a certain number of times, without any adjectives, then it means you do so independently.

Also, when you use this last phrase 'independently', it means you are being instructed on how to assign probabilities to the outcomes.