MEETING 20 - COMBINATORICS, PROBABILITY AND THE THE BINOMIAL THEOREM

Repetitions

We will study processes which involve selections in multiple steps. This will shed more light on in how many outcomes certain processes can have which will enable us to do more calculations on probability and understand the structure of sets better. At the foundation of these developments will be the multiplication rule. We study this by example and generalize into theorems.

The situation where we calculated the number of ways to put r *different* objects into n different boxes *with at most one object in every box* gave rise to the number

$$P(n, r) = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

A variation can be reached if all the objects are exactly the same, then the number of ways to put an object into r different boxes is the same number of ways in which we can choose r boxes out of n possible, and into those r boxes goes one object of the fixed kind. This number was found to be

$$\binom{n}{r} = \frac{n!}{r!(n-r)} = \frac{1}{r!}P(n, r).$$

Now we will study the corresponding situation where we allow *several* objects in each box, that is the meaning of the word *repetitions* in the header of this section. Let us do an example.

Example: In how many way can we put three marbles, one blue, one red and one yellow into 4 different containers if we can put any number of marbles in each box?

Solution: According to the multiplication rule, the red one can go into any of the four boxes, this can be done in 4 ways. But the same applies for the blue one, *irrespective* of that we put the red one in just before, so it is *4 ways again*, and the same for the last marble too, the total number of ways to place the marbles will be $4 \cdot 4 \cdot 4 = 4^3 = 64$ ways.

The above example illustrates the number of ways to put r marbles into n boxes, with any number of marbles being put into any box, when all the marbles are different. How about when all the marbles are the same? We will study this example but put another wording on it, we will give an example related to electrical engineering.

Example: Nicola Tesla, was a famous inventor who gave us the three-phase electrical power distribution system. Mr Tesla enjoyed feeding pigeons in the park and he reportedly took very great care to see to it that the pigeons got very good seeds. If 5 seeds are going to be distributed among 3 pigeons, in how many ways can this be done?

Solution: We will find this number by representing the situation in a smart way. We think of the distribution of 5 seeds among 3 pigeons. We represent such a distribution by a string of 5 zeroes (to symbolize the seeds) and 2 ones. First we put down a number of zeroes (possibly none!), this is the number of seeds that the first pigeon get. After this we put down a one. Thereafter follows yet another number of zeroes (possibly none!) and this symbolizes the number of seeds given to the second pigeon. After this we put down a one. Then the zeroes that remain will be given to the third pigeon. The strings will be 0000011, symbolizing that the first pigeon got all the seeds, 0000101 symbolizing that the first pigeon got 4 seeds, the second pigeon got one seed and the third pigeon got none, and so on until 1100000 which symbolizes that the third pigeon got all seeds. Since every string represents precisely one distribution of seeds among pigeons, to count the number of ways to distribute seeds as described, is equivalent to counting the numbers of strings with ones and zeroes of length 7 with precisely two ones. How many string are there of this kind? Well, we are counting the number of strings with a total of $5 + 2 = 7$ symbols, as soon as we placed the ones the other places will be filled with zeroes. How many ways are there to place those 2 ones? Well, we have figured this out earlier, it is a matter of selecting 2 objects out of 7 possible, so the correct answer is

$$\binom{7}{2} = 21 = \binom{3+5-1}{3-1} = \binom{n+r-1}{n-1} = \binom{n+r-1}{r} \text{ (since } \tbinom{p+q}{p} = \tbinom{p+q}{q} \text{ for all } p, q).$$

TABLE 1. The number of ways to put r marbles into n numbered boxes are

	Same marbles	Different marbles
At most one to a box	$\binom{n}{r}$	$P(n,r)$
Any number in a box	$\binom{n+r-1}{r}$	n^r

where $n = 3$ and $r = 5$. Now we have shifted to symbols n, r because there is nothing in this argument that is specific to 3 or 5, it works perfect in general and therefore the formula we have found must be valid for all n, r. We can list all the possible distributions of seeds among the pigeons:

$$0000011, 0000101, 0001001, 0010001, 0100001, 1000001, 0000110, 0001010, 0010010, 0100010, 1000010,$$

$$0001100, 0010100, 0100100, 1000100, 0011000, 0101000, 1001000, 0110000, 1010000, 1100000.$$

As we see there are 21 such strings. That this procedure is applicable to situations where we put identical marbles in boxes means that we have found another way of counting repetitions. We summarize the different methods in table (1).

We will end this section by looking at a particular application of the multiplication rule: re-arranging similar and non-similar objects. We will just take an example directly.

Example: The word ANNAGRAM is a mis-spelling of the word *anagram* and an anagram is a sequence of letters that are contained in a certain word. For example an anagram of the word THINK would be KHNIT. It does not have to be a word, just a sequence of letters. Now, how many anagrams does the mis-spelled word ANNAGRAM have?

Solution: We can solve this problem by using the multiplication rule and view the formation of an anagram as a multistep process and filling in the letters into positions, we have 8 positions to fill in and we do it like so:

1. Choose where to put the A's: we have 8 positions and 3 A's so this can be done in $\binom{8}{3}$ ways.
2. Choose where to put the N's: we have 5 positions left and 2 N's so this can be done in $\binom{5}{2}$ ways.
3. Choose the constellation of the letters G, R, and M: we have 3 positions left and placing the three letters G,R,M at those positions can be done in 3! ways.

In total there are $\binom{8}{3} \cdot \binom{5}{2} \cdot 3! = \frac{8!}{5!3!} \frac{5!}{3!2!} 3! = \frac{8!}{3!2!} = 4 \cdot 7 \cdot 6 \cdot 5 \cdot 4 = 3360.$

ELEMENTARY PROBABILITY

Probability is an *extremely* important branch of applied mathematics that has led to applications from which we can create forecasts that support the design of modern society. Insurance companies rely on these applications as do lottery vendors and linguists who study how human language is formed. *Many* other applications could be mentioned and the field that is developed from this is called *statistics* or *mathematical statistics* and it is of significant importance in society today.

Most people have a notion of what probability means just by a sense of common mathematical intuition, we can say things like "it will be about 16% chance to get 1 when we roll a fair die" or "it is not so probable that lightning strikes at the same place twice". Some people have misconceptions about probability too though, it is not so likely for an airplane to be subject to having two bombs on board at the same time, this is true, but this will not mean that it becomes safer to *bring along a bomb* on the airplane so as to guarantee the low probability of there being another one onboard. We will do some examples before we start with formal definitions.

Example: Two fair coins are tossed, give the probability that (a) both will come up heads, (b) they will have different outcomes, (c) they will have the same outcome.

Solution: (a) Toss first one coin. It is 50% that this will be heads. So in about 50% of all the cases in which we toss a fair coin, we will get heads. Now we toss the other, again it will be 50% probability that we will get heads, so that to get two heads will occur in 50% *of* 50% of all cases, this means that the probability will be $(50\%) \cdot (50\%) = 1/2 \cdot 1/2 = 1/4 = 25\%$. The probability of getting two heads is therefore 25%. (b) Independent of whether the first coins outcome is heads or tails, it will be 50% chance that the second coin that is tossed has an outcome different from the first. So we will get the probability being $(100\%) \cdot (50\%) = 50\%$. The exact same resoning applies to (c).

We will introduce some terminology here to be able to pin down exactly what we mean by probability and related concepts:

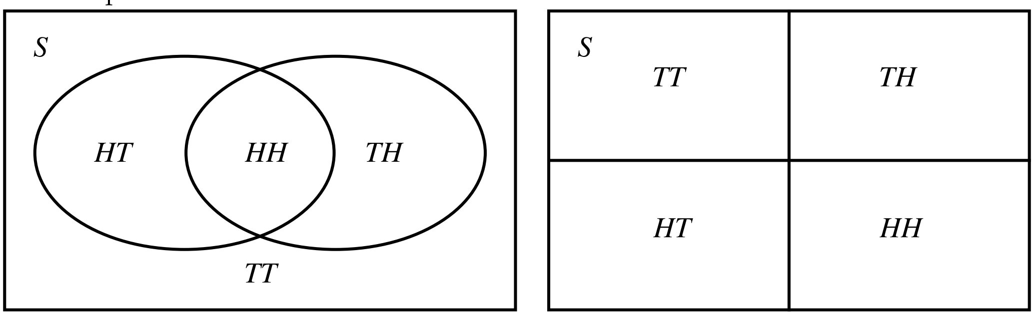
Definition: The word *experiment* is used for a procedure (like tossing coins or rolling dice) which has a possible set of outcomes, and this set of outcomes is called the *sample space* of the experiment. (The sample space of tossing two coins would be $\{heads - heads, heads - tails, tails - heads, tails - tails\}$). The word *event* is used for a subset A of the sample space. (The event above, called "both coins have different outcome" would then be the subset $\{heads - tails, tails - heads\}$ which of course is a subset of $\{heads - heads, heads - tails, tails - heads, tails - tails\}$). To achieve a more compact notation we will write HH for $heads - heads$, HT for $heads - tails$ and so on.

So probability is a measure of how likely a certain event is. This means that we are assigning a real number between 0 and 1 to a set, a probability is therefore a real-valued function $P : S \to [0, 1]$, where S is the sample space of some describable experiment or procedure. The function P has certain other demands on it that we will mention as we go along but we will wait with the formal definition (and the other demands on the function P). We will merely note that we have found

$$P(\text{"coins have different outcome"}) = 0.5 \text{ (or 50\%, remember percent simply means "hundredths")}$$

$$P(\text{"both coins come up heads"}) = 0.25 \text{ (or 25\%)}$$

The sample space, S, is the space of all possible outcomes of the experiment that we are studying. Since events are subsets of the sample space the sample space is a universal set to all events. This means that we can illustrate problems and application probability theory with Venn diagrams to use our sense of set theory to also get a feel for what is going on in probability theory. For the experiment with tossing two coins we can draw two non-similar-looking, but actually essentially identical, Venn diagrams that illustrate various aspects of the experiment:



Remember that the sample space S only consists of four elements, $S = \{HH, HT, TH, TT\}$. In both diagrams we have depicted the positions of these four outcomes. We will concern ourselves with the probability of two different events. The first event we will define as

$$A = \textit{The two coins have different outcomes or both coins come up heads}$$

and the second event we will define as

$$B = \textit{The two coins have the same outcome or both coins come up heads}.$$

These are sets and are defined with the logical *or*, as such, they will be unions of sets. If we do a little thinking we actually have the following set relationships

$$A = \{HT, TH\} \cup \{HH\}, \text{ and } B = \{HH, TT\} \cup \{HH\}$$

and from these relations we conclude that we can write the event B as $\{HH, TT\}$. Now when we compute the probabilities of the events A and B, since A is the disjoint union of $\{HT, TH\}$ and $\{HH\}$ we wish the probability of A to be the sum of the probabilities of $\{HT, TH\}$ and $\{HH\}$ respectively, that is we wish to have

$$P(A) = P(\{HT, TH\}) + P(\{HH\}).$$

As we have calculated these probabilities to 0.5 and 0.25 earlier, the probability of A becomes $P(A) = 0.75$, that is we have a 75% chance of obtaining different outcomes or both heads when we toss two coins. It is however a different story when we calculate the probability of B. As B is actually the set $\{HH, TT\}$, the event $\{HH\}$ is contained in it so that the event $\{HH\}$ will not increase the probability as it did with the event A. Now another interesting property of the probability will be used. Since S itself is a subset of S, this is also an event. This event is called the *certain* event and this will always happen, it is therefore assigned a probability of 100%, or 1. Now for any event E we have E^c also being an event and we always have, for events, $E \cup E^c = S$. This means that

$$P(E \cup E^c) = P(S) = 1 \Leftrightarrow P(E) + P(E^c) = 1 \Leftrightarrow P(E) = 1 - P(E^c)$$

and as this applies for all events we must have the calculation $P(B) = P(\{HH, TT\}) = P(\{HT, TH\}^c) = 1 - P(\{HT, TH\}) = 1 - 0.5 = 0.5$, so we have found that the probability of the event B is 50%. This is a *very* common technique, to calculate the probability of the *opposite* or *complement* event. All this rests on the fundamental property of additivity of probabilities, for any two disjoint events A, B we have

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

We are now more ready to give a formal definition of what a probability is. We will make this in two definitions.

Definition: A finite set $S = \{x_1, x_2, \ldots, x_n\}$, where $\{x_i\}_{i=1}^n$ are outcomes of an experiment, is called a *finite sample space* if:

1. To each outcome x_i there is an associated number between 0 and 1 called an *individual probability*, which is denoted $p(x_i)$.
2. The sum of all individual probabilities in S is 1, that is $\sum_{i=1}^n p(x_i) = 1$.

This first definition lays probability of individual outcomes. To make it useful we must pass onto sets, we do this in our next definition:

Definition: Let S be a finite sample space of a given experiment or procedure with an associated individual probability p defined on it's outcomes. A *probability* P on S is then a real-valued function on the set of events (subsets) of S such that $P(A) = \sum_{x_i \in A} p(x_i)$.

With this definition of probability we will have the following:

Proposition: Let $S = \{x_1, x_2, \ldots, x_n\}$ be a finite sample space with a probability P. Then

1. For all events $A, B \subseteq S$: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
2. For any event $A \subseteq S$: $P(A) \leq 1$, $P(\emptyset) = 0$, and $P(S) = 1$.
3. For any event $A \subseteq S$: $P(A^c) = 1 - P(A)$.

Proof: We will first establish (1) for disjoint sets, assume therefore that $A = \{a_1, a_2, \ldots, a_j\}$ and $B = \{b_1, b_2, \ldots, b_k\}$ and that these set (events) are disjoint, they contain no common element. But then we clearly have

$$P(A \cup B) = \sum_{x \in A \cup B} p(x) = \sum_{x \in A} p(x) + \sum_{x \in B} p(x) = P(A) + P(B).$$

Now assume that the events (sets) A, B are not disjoint. Since we have the equality $A \cup B = A \cup (B - A)$ and since A and $B - A$ are disjoint sets, we must have $P(A \cup B) = P(A \cup (B - A)) = P(A) + P(B - A)$. But since B can be written as $B = (B - A) \cup (A \cap B)$ and the sets $B - A, A \cap B$ are disjoint we can apply (1) for disjoint sets so that we have $P(B) = P(B - A) + P(A \cap B) \Leftrightarrow P(B - A) = P(B) - P(A \cap B)$ and this equality combined with $P(A \cup B) = P(A) + P(B - A)$ gives (1). To prove (2) we observe that the empty event $\emptyset$ has, by definition, $P(\emptyset) = \sum_{x \in \emptyset} p(x)$ which is a sum with no terms which is 0. Further, any event A also has $P(A) = \sum_{x \in A} p(x) \leq \sum_{x \in S} p(x) = 1$ and this applies to S itself so that (2) is completely proven. To see that (3) also is true we need only to observe that $A \cup A^c = S$ and that A and A^c are disjoint events so that $P(A) + P(A^c) = P(S) = 1$ which yields (3). The proof is complete.

By induction we can extend (1) of the previous proposition to something that also relates to the principle of inclusion and exculsion so that we have, for any sequence of events $A_1, A_2, \ldots, A_n$:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n),$$

and, as a corollary, if all the events are pairwise disjoint we obtain

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^n P(A_i).$$

If the events $A_1, A_2, \ldots, A_n$ are all pairwise disjoint, they are called *mutually exclusive*.

Probability can just be viewed as set theory with a weight (probability) attached to each set and that this weight is additive. A probability is a *measure* of how large a set is and the largest set is S itself and it's size (weight) is 1. The smallest set is $\emptyset$ and it has the size (weight) 0. To illustrate this, we can revisit the example from last lecture where we counted the number of integers in $\{1, \ldots, 10000\}$ that were not divisible by 2, 3 or 5, in that example we defined $A = \{x \in \Omega; 2|x\}$, $B = \{x \in \Omega; 3|x\}$ and $C = \{x \in \Omega; 5|x\}$. And we found that

$$|A| = 5000, \quad |B| = 3333, \quad |C| = 2000, \quad |A \cap B| = 1666, \quad |B \cap C| = 666, \quad |B \cap C| = 1000, \quad |A \cap B \cap C| = 333$$

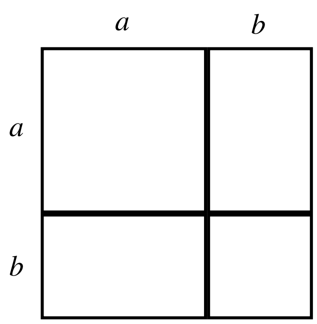
which by the principle of inclusion and exclusion allowed us to draw the conclusion that the number of integers in $\{1, 2, \ldots, 10000\}$ not divisible by 2, 3, or 5 was $10000 - k = 10000 - 7334 = 2666$. We could just as well say that the probability that an integer in $\{1, 2, \ldots, 10000\}$ is not divisible by 2, 3 or 5 is $2666/10000 = 26.66\%$. The thing is that we could have reached this conclusion by applying the probability version of the principle of inclusion and exclusion by stating that A is the event that an integer in $\{1, 2, \ldots, 10000\}$ is divisible by 2, B is the event that an integer in $\{1, 2, \ldots, 10000\}$ is divisible by 3, and C is the event that an integer in $\{1, 2, \ldots, 10000\}$ is divisible by 5, and that we through our investigations also conclude $P(A) = 0.5$, $P(B) = 0.3333$, $P(C) = 0.2$, $P(A \cap B) = 0.1666$, $P(B \cap C) = 0.666$, $P(B \cap C) = 0.1$, and $P(A \cap B \cap C) = 0.333$. Now to find the probability that an integer in $\{1, 2, \ldots, 10000\}$ *is* divisible by 2, 3 or 5 we form the number

$$P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

and find that it is 0.7334. To find the probability that a number is *not* divisible by 2, 3 or 5 we just form $1 - 0.7334$ which is again 0.2666, that is 26.66%. All these steps in finding this probability corresponds exactly to the steps in which we found the number of integers in $\{1, 2, \ldots, 10000\}$ not divisible by 2, 3 or 5, so *elementary* probability is just set theory with a numerical value attached to each set and the usual set operations correspond to operations on the corresponding numerical values attached to these sets with union roughly meaning addition (forming $P(A) + P(B)$) and taking complement roughly meaning inversion (forming $1 - P(A)$).

The biniomal theorem

Consider a multiplication of the expression $(a + b)(a + b)$. We may remember that this multiplication turns out to be $a^2 + 2ab + b^2$. But how is this expression found? We can illustrate it with a figure:



When we calculate the expression $(a + b)(a + b)$ it can be likened to calculating the area of a square whose side is $a+b$, now when multiplying the sides with each other (as one does when calculating the area of a square) we need to multiply each term in $(a + b)$ (the first side) with each term in $(a + b)$ (the other side), this means forming the expressions $a \cdot a = a^2$, $a \cdot b = ab$, $b \cdot a = ab$, and $b \cdot b = b^2$. This can be thought of, combinatorially, to choose every term in $(a+b)$ (there is only 2 of them) and multiplying them with every term in $(a+b)$ (only 2 here also), combinatorially this makes $2 \cdot 2 = 4$ terms whose expressions we have listed. Summing upp all these terms we get $a^2+ab+ab+b^2 = a^2+2ab+b^2$, the well-known expansion of $(a+b)^2$. Let us generalize this process to higher powers than 2. The exact same procedure works for the power 3: $(a + b)^3 = (a + b)(a + b)(a + b)$, and when we calculate this we can reason in the same way: each term in $(a + b)$ (the first factor), must be multiplied by each term in $(a + b)$ (the second factor) and this in turn must be multiplied by each term in $(a + b)$ (the last factor). This makes up for a total of 8 terms, they are $aaa, aab, aba, baa, abb, bab, bba, bbb$. Of course we have $aab = aba = baa = a^2b$ and $abb = bab = bba = ab^2$ and gathering these together we get 3 of these and gathering all terms with matching powers of a and b we get the computation

$$(a + b)(a + b)(a + b) = aaa + aab + aba + baa + abb + bab + bba + bbb = a^3 + 3a^2b + 3ab^2 + b^3.$$

We see that we are gathering terms together which have the same powers of a and b, and every term must appear since this is the way we compute an expression like $(a + b)^3$. The question is what are the coefficients in front of a term expressed like $a^{something}b^{someting}$ in the expression of the type $(a + b)^n$? This is what the binomial theorem states. To begin with the expansion of $(a + b)^n$ must have terms of the sort a^ib^k where

$i+k=n$. Above we saw that $(a+b)^3$ had terms of the sort $a^3=a^3b^0$, and $a^2b=a^2b^1$ and so on. We can therefore write the expansion of $(a+b)^n$ as a sum of terms like this:

$$(a+b)^n = (\text{coeff})\cdot a^n b^0 + (\text{coeff})\cdot a^{n-1}b^1 + \ldots + (\text{coeff})a^1b^{n-1} + (\text{coeff})\cdot a^0 b^n \tag{1}$$

the problem is to understand what the coefficients (denoted by the marker (coeff) in the expression) are. The beauty of the binomial theorem is that we can understand what they are by considering what $(a+b)^n$ really is. It is the product of n factors like this:

$$(a+b)\cdot(a+b)\cdot\ldots\cdot(a+b)$$

and to expand this, each term (a or b) in each factor $(a+b)$ must be multiplied with every other term in all the other factors $(a+b)$. This gives rise to an expression like (**??**), and coefficient in front of a^n will be 1, since there is only 1 way to get a^n when selecting terms (a or b's) in the manner described (one from each factor $(a+b)$). Similarly the coefficient in front of $a^{n-1}b$ must be $\binom{n}{1}=n$ since this the number of ways in which we can choose one term b and all other terms are chosen to a from the n factors $(a+b)(a+b)\ldots(a+b)$. This goes on and on, the coefficient of $a^{n-2}b^2$ must be $\binom{n}{2}$ since this is the number of ways to choose 2 parentheses out of n in which we choose the term b. In conclusion, and this is the binomial theorem:

The Binomial Theorem: For each natural number n and every pair of complex numbers a, b we have

$$(a+b)^n = \binom{n}{0}a^nb^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n-1}a^1b^{n-1} + \binom{n}{n}a^0b^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k.$$

The expression $a+b$ is called a *binomial* (*bi* for "two", there are 2 terms in it, a and b) and therefore the expressions $\binom{n}{k}$ are called *binomial coefficients*. In the reasoning above (which could be called a proof) we saw how combinatorics (reasoning out how many ways something can be done) ties in beautifully with this algebraic result (which is called the Binomial Theorem).

Another very nice illustration of how tied together with combinatorics the binomial coefficients are is illustrated by the following property that relies on the addition rule: The binomial coefficient $\binom{10}{6}$ is the number of ways in which we can choose 6 objects from a set of 10. Denote these 10 objects by $o_1, o_2, \ldots, o_{10}$. There are two distinct ways to choose 6 objects from these 10, either o_1 is among the chosen objects or o_1 is not among the chosen objects. So

$$\binom{10}{6} = \textit{Number of ways to choose 6 from 10, not choosing } o_1 + \textit{Number of ways to choose 6 from 10, choosing } o_1.$$

But the *Number of ways to choose 6 from 10, not choosing* o_1 must be $\binom{9}{6}$, since we are choosing all 6 objects from the 9 objects $o_2, o_3, \ldots, o_{10}$, and the *Number of ways to choose 6 from 10, choosing* o_1 must be $\binom{9}{5}$ since once we have decided to choose o_1, it remains to choose the last 5 again from the 9 objects $o_2, o_3, \ldots, o_{10}$. So we have the general formula

$$\binom{10}{6} = \binom{9}{6} + \binom{9}{5} \qquad \text{or, since this can be done for any } n, k \text{ (with } k<n\text{):} \qquad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

This last identity gives rise to a structure on all binomial coefficients, we can arrange them in what is known as *Pascal's Triangle*. Since we always have $\binom{n}{0}=\binom{n}{n}=1$ for all natural numbers n, and the formula above holds, we can arrange the binomial coefficients like this:

$n=0$:					1				
$n=1$:				1		1			
$n=2$:			1		2		1		
$n=3$:		1		3		3		1	
$n=4$:	1		4		6		4		1

The upper edges are $\binom{n}{0}=\binom{n}{n}=1$, those are all the 1's. The interior of the triangle are the binomial coefficients which are computed by the formula given, for example $\binom{2}{1}=\binom{1}{0}+\binom{0}{0}=1+1=2$, $\binom{3}{1}=\binom{3}{0}+\binom{2}{1}=1+2=3$, and so on. Indeed, we can see that the coefficients of the expansions of the expressions $(a+b)^2$, $(a+b)^3$, and $(a+b)^4$ are embedded in the triangular structure, they are each on row 2, 3 and 4 respectively.

Example: Does the expansion of $\left(2x^2-\frac{1}{x}\right)^{10}$ have any constant term? Any x-term? Any x^2-term?

Solution: A straight-forward application of the binomial theorem gives

$$\left(2x^2 - \frac{1}{x}\right)^{10} = \sum_{k=0}^{10} \binom{10}{k} (2x^2)^{10-k} \cdot (-\frac{1}{x})^k = \sum_{k=0}^{10} \binom{10}{k} 2^{10-k} x^{20-2k} (-1)^k x^{-k} = \sum_{k=0}^{10} \binom{10}{k} 2^{10-k} (-1)^k x^{20-3k}.$$

We are asking the question if this expression can have a constant term? That would require the exponent of x to be 0, can it be 0? No since it is of the for $20 - 3k$ and if this would be 0 for a certain k, then 20 would be divisible by 3 which it is not. Hence the there is no constant term, or rather it is 0. Can there be any x-term? We then ask ourselves if there can be any term which has the exponent of x being 1, that is can $20 - 3k$ be 1 for some k? Again the answer is no since we would then have $19 = 3k$ and then 129 would be divisible by 3, which it is not. (Since 19 is even a prime number.) The situation is different with the x^2-term, yes we can have a non-vanishing x^2-term and the coefficient for the x^2-term will be the expression

$$\binom{10}{k} 2^{10-k} (-1)^k$$

for that k which makes the exponent of x equal to 2, this will be the k that solves $20 - 3k = 6$ which is $k = 6$. And so the coefficient of the x^2-term is

$$\binom{10}{6} 2^{10-6} (-1)^6 = 2^4 \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2} = 2^4 \cdot 10 \cdot 3 \cdot 7 = 210 \cdot 16 = 3360.$$