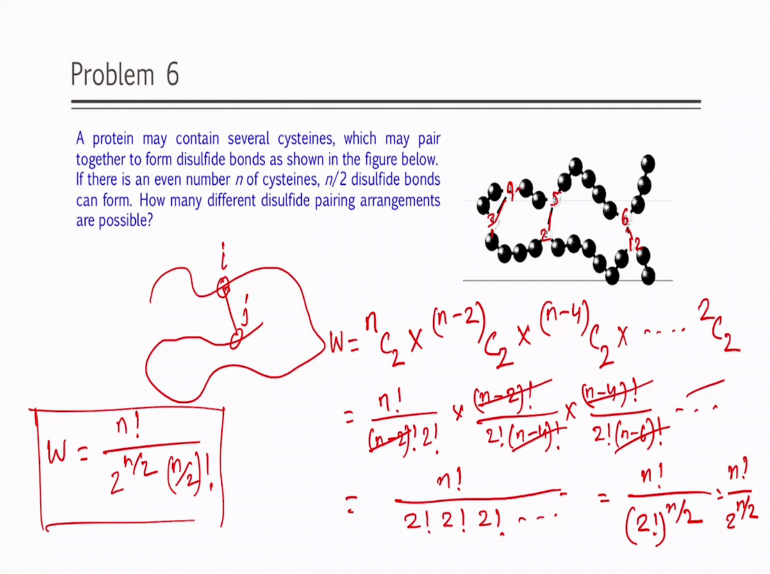
Chemical Principles II
Professor Dr Arnab Mukherjee
Department of Chemistry
Indian Institutes of Science Education and Research, Pune
Tutorial Problem – 11

Ok so we will continue with the tutorial problems on molecular thermodynamics part. First we will talk about the problem of bond formation in a protein, so the question is like this. A protein may contain several cysteines which may pair together to form die sulphide bonds as shown in the figure below you know means side.

(Refer Slide Time: 0:37)



Problem 6

A protein may contain several cysteines, which may pair together to form disulfide bonds as shown in the figure below. If there is an even number *n* of cysteines, *n*/2 disulfide bonds can form. How many different disulfide pairing arrangements are possible?

$$W = {}^{n}C_2 \times {}^{(n-2)}C_2 \times {}^{(n-4)}C_2 \times \cdots {}^{2}C_2$$

$$= \frac{n!}{\cancel{(n-2)!}\,2!} \times \frac{\cancel{(n-2)!}}{2!\,\cancel{(n-4)!}} \times \frac{\cancel{(n-4)!}}{2!\,\cancel{(n-6)!}} \cdots$$

$$= \frac{n!}{2!\,2!\,2!\cdots} = \frac{n!}{(2!)^{n/2}} \doteq \frac{n!}{2^{n/2}}$$

$$\boxed{W = \frac{n!}{2^{n/2}\,(n/2)!}}$$

If there is an even number n of cysteines, n by 2 disulphide bonds can form, how many different disulphide pairing arrangements are possible? So you know the cycteine is cysteine has S as a terminal group and that S can form bond with another sulphur group and therefore cysteines can form disulphide bonds and these bonds make a particular confirmation of the protein more stable. So here the question however is much more theoretical in nature, so if there are n cysteines groups we know that there will be n by 2 bonds because 2 system groups will form a bond and how many such bonds or pairing are possible.

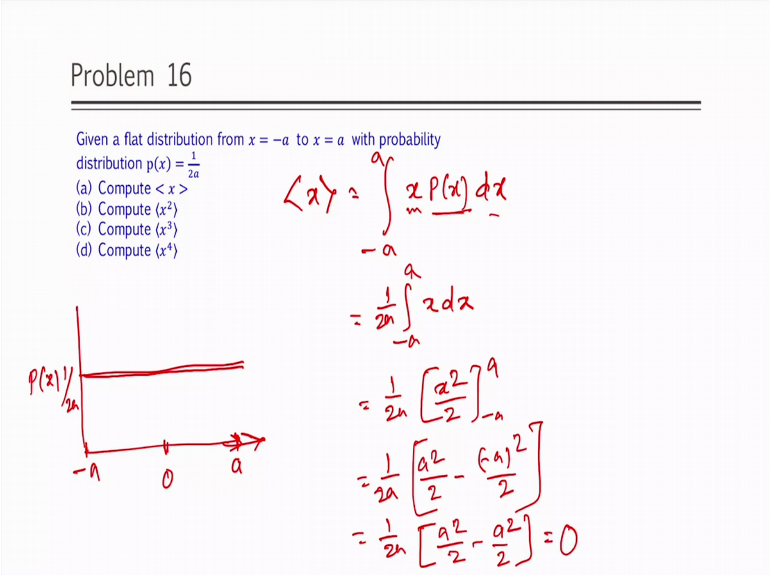
Now those pairing will be unique you know indistinguishable pairing that you have to remember, this picture may not be very clear so it is like this that so this is the 3rd cysteine, this is the 4th cysteine, they are forming a bond, 5th and 2nd are forming bond, and 6th and maybe 12th are forming bond, so these are just examples. So we can also think of the problem as random chain of protein where we have certain pairs of I and J system group that are forming bond and then we pick among the rest of the N – 2, two more atoms and things like that.

So how you can think about that is that, out of the N cysteine we can take any 2 cysteine to form a bond in N say 2 number of ways that will leave us with N – 2 cysteine residues, again we can take 2 out of them to form a bond that will leave us with N – 4 cysteine, again we can take 2 and like that you can continue and then last two will form the bond in two ways that means unique way, so that will be the number of possible arrangements. Now if I write it down, it will be n factorial by n - 2 factorial and 2 factorial followed by n - 2 factorial by 2 factorial n - 4 factorial followed by n - 4 factorial by 2 factorial n - 6 factorial and it will continue.

So this n - 2 factorial and n - 2 factorial will cancel, n – 4 and n - 4 will cancel and what is going to happen as you can see that we will end up with n factorial by 2 factorial 2 factorial 2 factorial and things like that. So how many towns will be there? As you can see that we are going to make n by 2 pairing so there will be n by 2 terms so therefore n factorial by 2 factorial to the power n by 2, which is n factorial by 2 to the power n by 2, so these many bonds will form but these all bonds are unique bonds right now.

So since we have no preference of any of those n by 2 pairing, the number of pairings possible will be n factorial by 2 to the power n by 2 divided by n by 2 factorial because n by 2 factorial we are doing so that we are saying that all n by 2 pairings are indistinguishable which means that 3-4 and 2-5 are not distinguishable because these are just the pairings, they are just bonds so we are not distinguishing bond between 3 and 4th amino acid versus 2nd and 5th amino acid versus 6th and 12th amino acid, so these are the number of possible ways the pairing can be possible for N cysteines where 2 cysteines form a double bond.

(Refer Slide Time: 6:59)



Problem 16

Given a flat distribution from $x = -a$ to $x = a$ with probability distribution $p(x) = \frac{1}{2a}$

(a) Compute $\langle x \rangle$
(b) Compute $\langle x^2 \rangle$
(c) Compute $\langle x^3 \rangle$
(d) Compute $\langle x^4 \rangle$

$$\langle x \rangle = \int_{-a}^{a} x\, p(x)\, dx$$

$$= \frac{1}{2a}\int_{-a}^{a} x\, dx$$

$$= \frac{1}{2a}\left[\frac{x^2}{2}\right]_{-a}^{a}$$

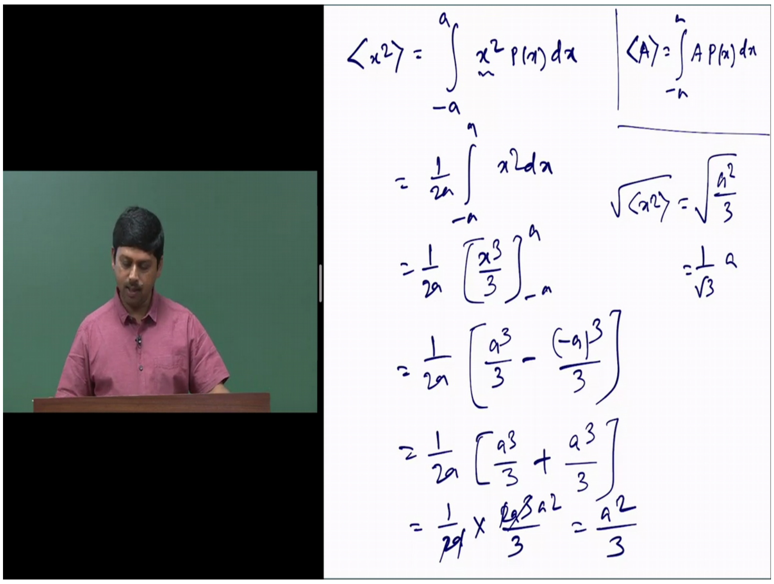
$$= \frac{1}{2a}\left[\frac{a^2}{2} - \frac{(-a)^2}{2}\right]$$

$$= \frac{1}{2a}\left[\frac{a^2}{2} - \frac{a^2}{2}\right] = 0$$

Okay the next problem, given a flat distribution fromX equal to - A to X equal to + A with a probability distribution P X as 1 by 2 A, compute these quantities. So these are practise problems for you since you know that averaging is very important statistical thermodynamics or molecular thermodynamics, money observable you obtain is nothing but an average value weighted by their own probabilities. So here we are talking about a flat distribution which means that the probability is same everywhere in the range of range between - A to + A, it is same everywhere and that value is nothing but 1 by 2 A so that will look something like that.

And this kind of situation arises as you know for cysteine with all eco-energy levels where there is no preference between one energy level versus another one, so this is a situation something like that, once you understand that once you understand how to calculate the average, you will be able to do that for other systems as well.

(Refer Slide Time: 8:20)



$$\langle x^2\rangle = \int_{-a}^{a} x^2 P(x)\,dx$$

$$\langle A\rangle = \int_{-a}^{a} A\,P(x)\,dx$$

$$= \frac{1}{2a}\int_{-a}^{a} x^2\,dx$$

$$= \frac{1}{2a}\left[\frac{x^3}{3}\right]_{-a}^{a}$$

$$= \frac{1}{2a}\left[\frac{a^3}{3} - \frac{(-a)^3}{3}\right]$$

$$= \frac{1}{2a}\left[\frac{a^3}{3} + \frac{a^3}{3}\right]$$

$$= \frac{1}{\cancel{2a}} \times \frac{\cancel{2a^3}\,a^2}{3} = \frac{a^2}{3}$$

$$\sqrt{\langle x^2\rangle} = \sqrt{\frac{a^2}{3}}$$

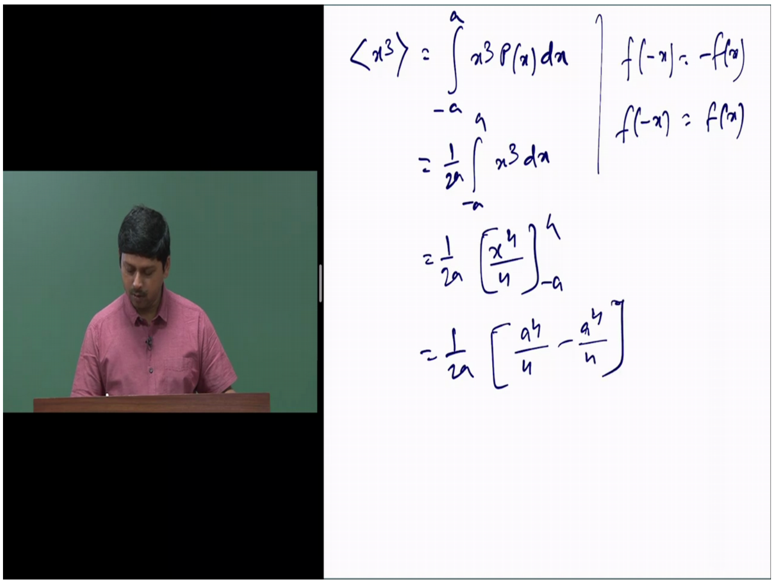
$$= \frac{1}{\sqrt{3}}\,a$$

So let us calculate the average X, so average X or any average value for that matter should be calculated either by sum or by integration, here we are going to use integration because here X can take any values so the range of X is from - A to + A and the average of X is nothing but X into P X dX so this is the formula. The reason is that any quantity that we are talking about for which we are going to take the average has debated by the probabilities of obtaining that and then d S is nothing but for integration sign so now it will give us - A to A and X is X dX and we know the value of PX which is 1 by 2 A and that is a constant quantity so it is 1 by 2 A X square by 2 going from - A to + A which will give us 1 by 2 A, A square by 2 - A square by 2 which is 1 by 2 A A square by 2 - A square by 2 giving us 0 so average value of X will give us 0.

So now I am going to do for X square average, so X square average is not X average square, you know that is that will be wrong so it is - A to A nothing but it is a square then P X and dX, so you see I have just replaced X by X square. Now just imagine I am going to take an average of some quantity A for which the probability is P X so it will be A P X dX, now A can be X or X square or X cube it does not matter ok. So we are going to do that, now this is again very simple 1 by 2 A - A to A X square dX, 1 by 2 A X cube by 3 – A to A which is A cube by 3 - - A cube by 3 1 by 2 A A cube by 3 - - A cube will give us + so we are going to get 1 by 2 A multiplied by 2 A cube by 3, 2 2 cancels so we are going to get A square by 3.

So just to tell you that X average is called mean and X square average is the variance, so here mean will be 0 because it is a flat distribution, it can be anywhere from – A to + A so mean is in the middle, where as standard deviation will give you a measure of how much it goes and how much it varies and it turns out that will be not full A but A square by 3. If you take the root, standard deviation of that will be the root of that so this is variance and standard deviation will be root over of X square average which is root over of A square by 3 which is 1 by root 3 into A, so it is not fully A but it is bit less than A when you calculate that average.

(Refer Slide Time: 9:23)



$$\langle x^3 \rangle = \int_{-a}^{a} x^3 P(x)\, dx$$

$$= \frac{1}{2a} \int_{-a}^{a} x^3\, dx$$

$$= \frac{1}{2a} \left[\frac{x^4}{4} \right]_{-a}^{a}$$

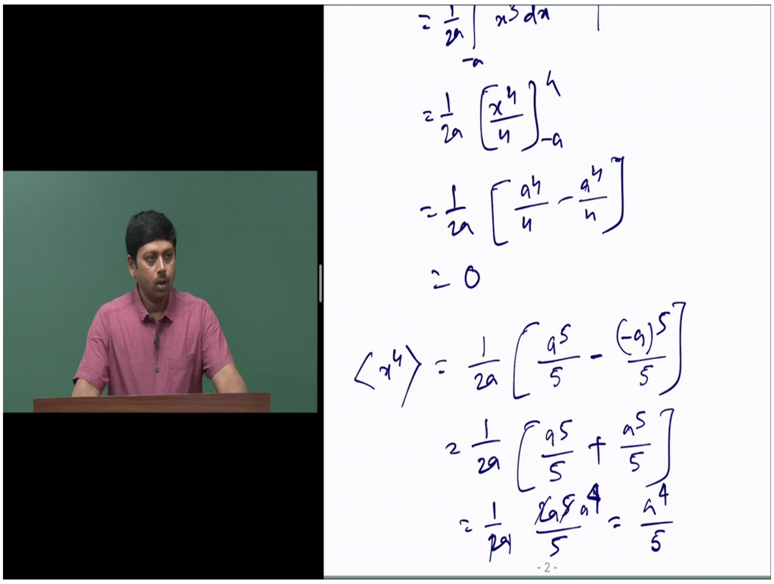
$$= \frac{1}{2a} \left[\frac{a^4}{4} - \frac{a^4}{4} \right]$$

$$f(-x) = -f(x)$$

$$f(-x) = f(x)$$

Okay so now we have to calculate X cube average and X 4 average, so without doing that everything for example when we talk about X cube average so it is again X cube P X dX, you have you should know by now that when you are talking about an odd function, odd functions are function where F of – X is - of F X, in those cases it will become 0 so you can see that right now. And for the even functions where F of - X equal to F of X that will not be 0, so we are going to do that again for X cube. So it will be 1 by 2 A – A to A X cube dX which will give us 1 by 2 A X to the power 4 by 4 going from - A 2 + A, 1 by 2 A, it will give us A to the power 4 by 4 – A to the power 4 b y 4 and it will going to give us 0.

(Refer Slide Time: 10:22)



$$= \frac{1}{2a}\int_{-a} x^3 dx$$

$$= \frac{1}{2a}\left[\frac{x^4}{4}\right]_{-a}^{a}$$

$$= \frac{1}{2a}\left[\frac{a^4}{4} - \frac{a^4}{4}\right]$$

$$= 0$$

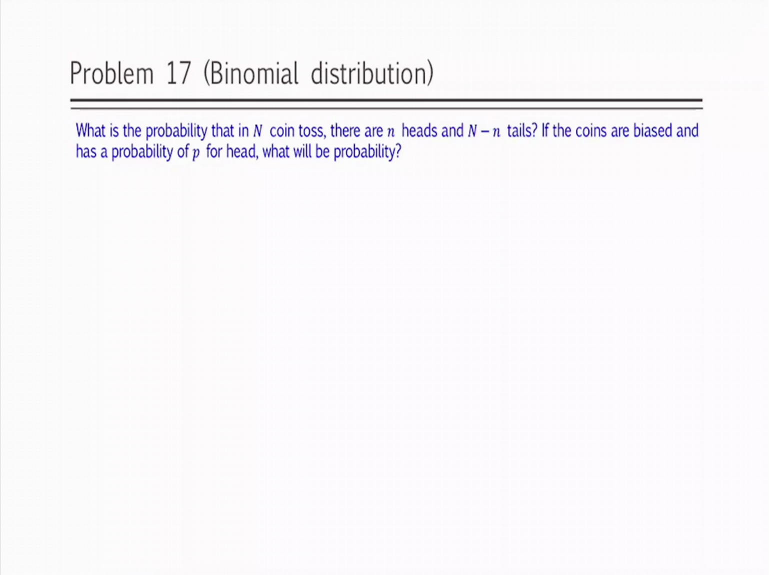
$$\langle x^4 \rangle = \frac{1}{2a}\left[\frac{a^5}{5} - \frac{(-a)^5}{5}\right]$$

$$= \frac{1}{2a}\left[\frac{a^5}{5} + \frac{a^5}{5}\right]$$

$$= \frac{1}{\cancel{2a}} \frac{\cancel{2a^5}\, a^4}{5} = \frac{a^4}{5}$$

If we do X to the power 4 that will not be the case, if I just write the answer I can write the answer now by following all the above ones is 1 by 2 A A to the power 5 by 5 - - A to the power 5 by 5 which is 1 by 2 A A to the power 5 by 5 + A to the power 5 by 5 which is 1 by 2 A 2 A to the power 5 by 5, it is going to give us A to the power 4 by 5 okay, so now we are going to go to the next problem.

(Refer Slide Time: 11:03)

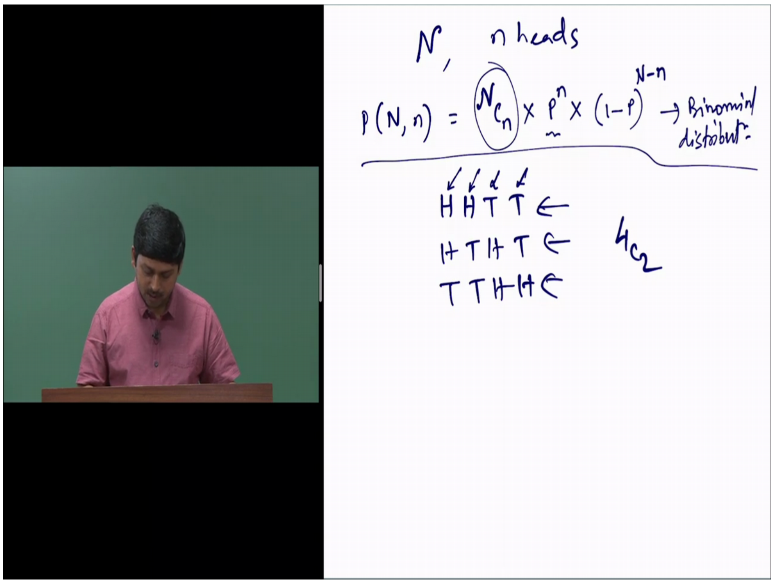


Problem 17 (Binomial distribution)

What is the probability that in N coin toss, there are n heads and $N-n$ tails? If the coins are biased and has a probability of p for head, what will be probability?

Ok so the next question is that what is the probability that in N coin toss there are n heads and N – n tails ? If the coins are biased and has a probability P for head, P can be anything other than half, if they are unbiased points then head and tail will have half probability right, if they are biased points then it can have P can have any values. For example, if P is 1 it will be always head and never tail, so what will be the probability. So this question will actually lead us to understand what is called binomial distribution, we have already talked about it but we are going to now discuss it in light of this particular problem today.

(Refer Slide Time: 12:04)

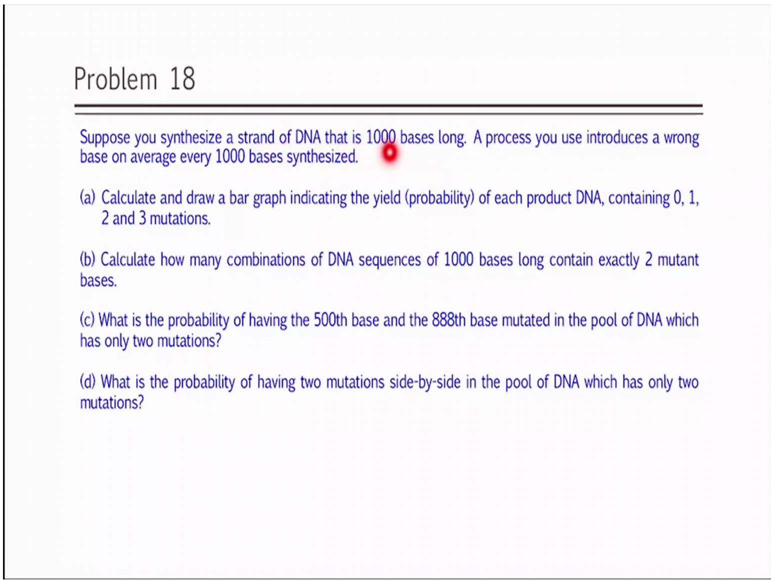


So N coin tosses and n heads, what is the probability and n heads then of course rest of them is tails right. So the probability is typically written as P N, n which says that how many ways I can first, let us calculate how many ways we can take we can get n heads out of N tail and that will be N C n that many ways you can do that 1st, I will explain that in a moment. And then we have n head with a probability P so it will be P to the power n that means the probability for tail is 1 – P and how many of them are there, N – n, so this is a probability of having n heads out of N coin tosses and this is called binomial distribution.

So this you know we know that for the example if we toss a coin and you can ask a question that in N coin tosses, what is the probability that all of them will be head? So that will be typically we say that it is half to the power n half half half half half however, here we are asking a slightly different question because we are asking that let us say 4 coin tosses, let us say what is the probability of 2 being head and 2 being tail so that scenario can be understood from so many different possibilities, it can have this situation will satisfy the thing, this situation will also satisfy the thing, this situation will also satisfy so therefore as you can see that out of 4 C 2 that means 6 possible ways we can have 2 heads and 2 tails. And since each of them are independent events, each of them will have probability half associated with it provided it is a biased coin then P and 1 - P are equal and therefore P is half.

However, if they are not biased then we have to associate the probability as P to the power N or here P to the power 2 and 1 - P to the power 2 as well. So and this is an important distribution as you know and the number of possibilities we have already discussed in case of Maxwell Boltzmann distribution that it is N C n. So now we are going to see one application of that which is in the next problem.

(Refer Slide Time: 14:49)



Problem 18

Suppose you synthesize a strand of DNA that is 1000 bases long. A process you use introduces a wrong base on average every 1000 bases synthesized.

(a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2 and 3 mutations.

(b) Calculate how many combinations of DNA sequences of 1000 bases long contain exactly 2 mutant bases.

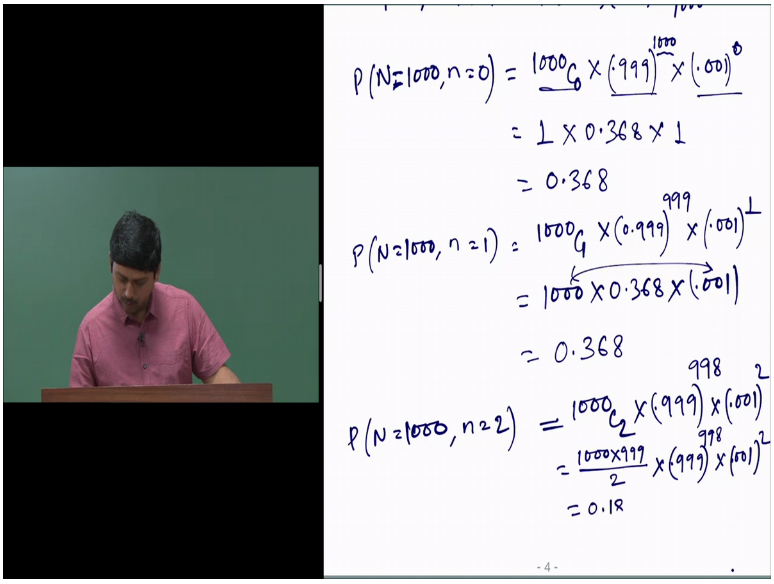
(c) What is the probability of having the 500th base and the 888th base mutated in the pool of DNA which has only two mutations?

(d) What is the probability of having two mutations side-by-side in the pool of DNA which has only two mutations?

So suppose you synthesize a strand of DNA that is 1000 bases long, a process you use introduces a wrong base on an average every 1000 bases fantasized, which means that if you synthesized 1000 base there is a probability that one of the base maybe wrong, so the probability of wrong base is 1 by 1000 and the probability of correct base is of course 999 by 1000. So keeping that in mind it is almost like this binomial distribution right because whenever there are 2 events head or tail or mutation or non-mutation, there are 2 outcomes so then the probability will be either P and 1 - P and the distribution of that will follow binomial distribution. So here we can think that probability of correct bases is something like in head and probability of incorrect bases is something like tail, so only thing is that P is 999 by 1000 and 1 - P is of course 0.001 or 1 by 1000.

So now the question is that calculate and draw a bar graph indicating the yield of each product DNA containing 0, 1, 2 and 3 mutations. So it is asking us to plot the binomial distribution where n in the earlier problem was 0 or 1 or 2 or 3 ok, so basically we are going to follow the same thing.

(Refer Slide Time: 16:28)



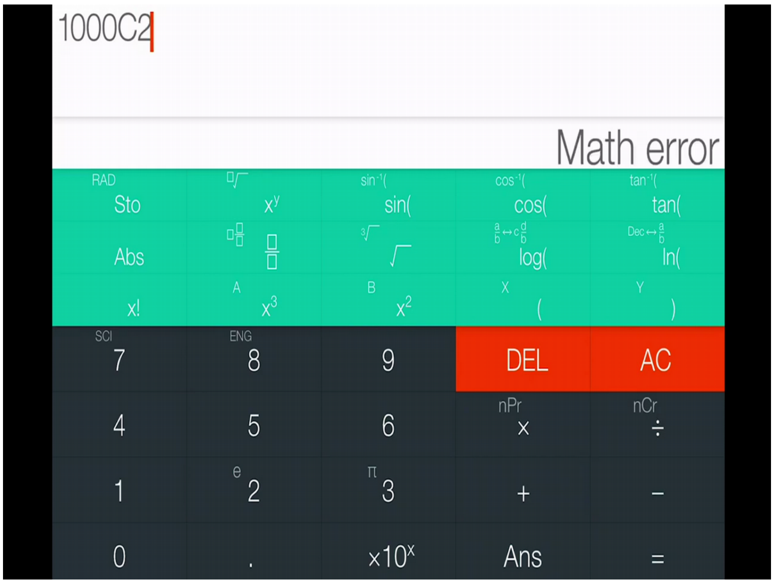
$$P(N=1000, n=0) = {}^{1000}C_0 \times (.999)^{1000} \times (.001)^0$$
$$= 1 \times 0.368 \times 1$$
$$= 0.368$$
$$P(N=1000, n=1) = {}^{1000}C_1 \times (0.999)^{999} \times (.001)^1$$
$$= 1000 \times 0.368 \times (.001)$$
$$= 0.368$$
$$P(N=1000, n=2) = {}^{1000}C_2 \times (.999)^{998} \times (.001)^2$$
$$= \frac{1000 \times 999}{2} \times (.999)^{998} \times (.001)^2$$
$$= 0.18$$

So here as you see that our head is equivalent to correct base with a probability P as 999 by 1000 which is 0.999, so out of 1000 bases 999 are on an average. Tail is correspond to mutated base or incorrect base so it is denoted as 1 - P which is 1 by 1000 or 0.001. Once you know that you have to calculate the probability of P and N where N equal to 1000 and n equal to 0, how many ways you can do that? Of course, it is 1000 C 0 P to the power N which is 0.999 to the power 1000 of course and 0.001 to the power 0. So 1000 C 0 is 1000 factorial by 1000 and 0 factorial which is just 1, and 0.999 to the power 1000 this one of course will be 1, this is of course is 1, this one you have to calculate so let us use the calculator and calculate that, 0.999 to the power 1000 is 0.368 into 1000 which is 0.368.

Let us do the next one, P N equal to 1000 again because out of 1000 bases we are going to consider n equal to 1, again that will be 1000 C 1 because mutation can happen at any point right, in the 1000 bases the mutation can happen in the 1st base or in the 2nd base or in the 3rd base, so there are 1000 possibilities permutations to happen and that is what 1000 C 1 indicates. Now 0.999 to the power so now this will be 999 and 0.001 to the power 1, so this is nothing but 1000 multiplied by 0.999 to the power 999 let us see how much it is… 0.368 again and 0.001. So now if you multiply this one with this one you are going to get one so you are left with almost the same number, now let us see the next one.

1000, n equal to 2 that is going to give us 1000 C 2 multiplied by 0.999 to the power 998 into 0.001 to the power 2. What is 1000 C 2, so let us do the complete calculation in the calculator 1000 C 2 into 999 raise to 998 and things like that. Ok 1000 C 2 we will do by hand because calculator may not work, 1000 into 999 divided by 2 and then 0.999 to the power 998 which will be almost 1368 and then 0.001 to the power 2, so let us do that.

(Refer Slide Time: 19:51)



1000C2

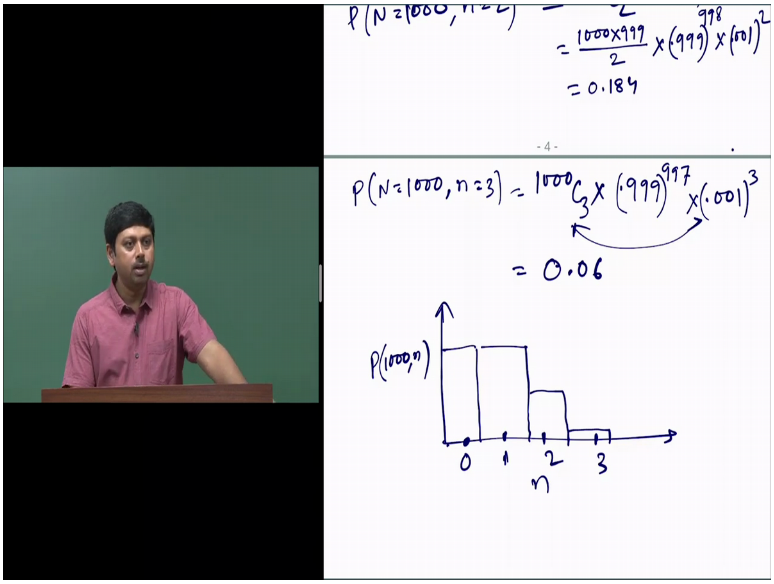
Math error

$.999^{998}\times(500\times999)\times.000001$

0.184031744129611633947350420030

So as you can see 1000, this calculator 1000 C 2 will give you error so we are going to do our own way, so 1000 into 999 divided by 2 okay, so that is reasonable and that is multiplied by 0.999 to the power 998 multiplied by 0.001 to the power 2 ok, so we will do step-by-step, let us see how much this gives us. Okay we will delete this one and see, okay as I said that will be 0.368 and so how much we will get, 1000 or we can just write 500 because we know that into 999 ok, so this much you got and 0.001 to the power 2 0.001 is 10 to the power - 3 so it will be 10 to the power – 6. And if I do 10 to the power - 6, it will become let us say we will do that 0.000001 so it will be 0.184 so I have just written down instead of doing it so 0.184, this is the value that we got.

(Refer Slide Time: 22:06)



$$= \frac{1000\times 999}{2}\times(.999)^{998}\times(.001)^{2}$$

$$= 0.184$$

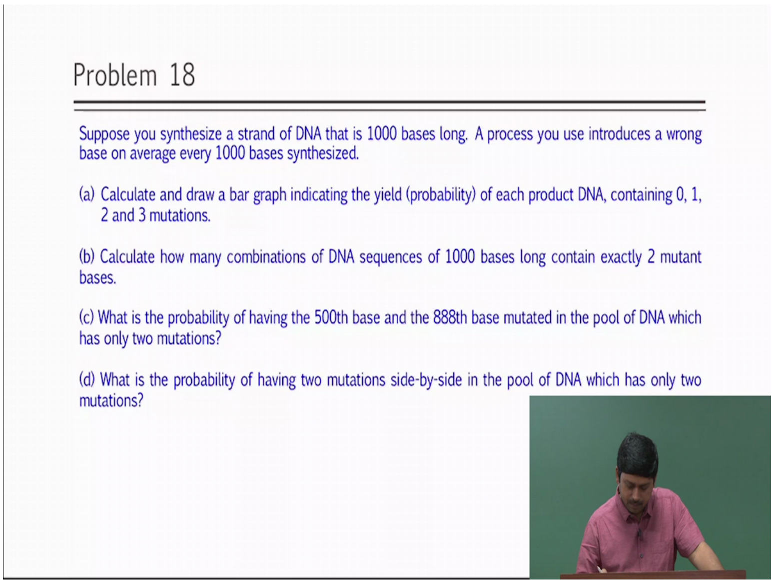
$$P(N=1000, n=3) = {}^{1000}C_3 \times (.999)^{997} \times (.001)^3$$

$$= 0.06$$

And now the last one which is N is equal to 1000, n is equal to 3. It will be 1000 C 3 into 0.999 to the power 997 into 0.001 to the power 3. The 1st we will do 1000 C 3, we will do one by one because otherwise it will not work, so 1000 C 3 is 1000 into 999 into 998 divided by 3 factorial which is 6, so that is this much ok. So this is followed by 0.001 to the power 3 ok good 0.166 and then we are still left with, so these 2 give 0.166 and then we are still left with 0.999 to the power 997 let us see, it will be almost 0.368 though ok we have got 0.06 as in the result 0.06. Now if you plot a bar diagram of those probabilities, P and N which is 1000, n and this is n where n is 0, 1, 2 and 3, so bar diagram for 0 it was 0.368 and this is also 0.368, this was 0.18 and this was 0.067 so the bar diagram will look something like this.

So as you can see that immediately when it goes to 3, you know the probability just drops like anything, if you go to 4 it will be like 0 so that means if the probability that 1 out of 1000 bases will be incorrect in a 1000 base sequence then the (())(24:15) will go to 4 or 5 is almost 0, it will not basically go to 0, so it just shows that how the probability is going to change with number of mutations given that. So it is like this like if you toss and you know that head and tail are equally likely then what is the probability that you are going to get like one after another tail or head, what is the probability that you get 5 heads in a row? It will be very very less so that is the thing.

(Refer Slide Time: 24:49)



Problem 18

Suppose you synthesize a strand of DNA that is 1000 bases long. A process you use introduces a wrong base on average every 1000 bases synthesized.

(a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2 and 3 mutations.

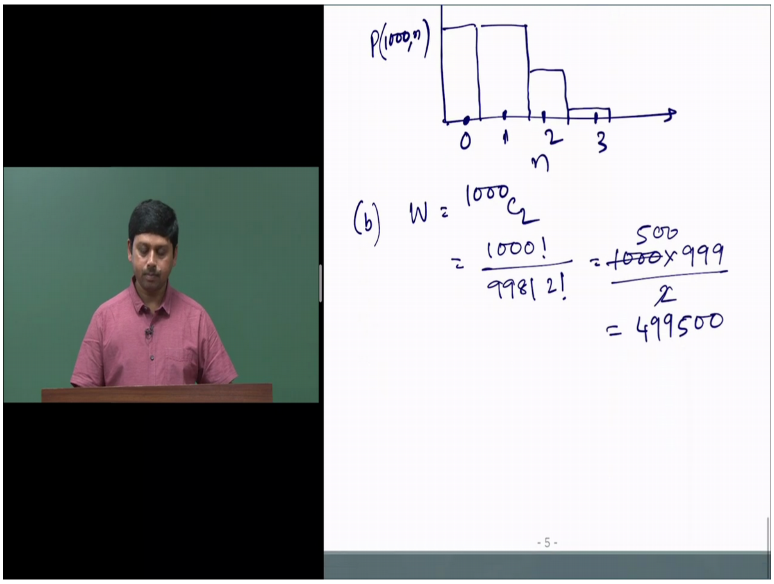
(b) Calculate how many combinations of DNA sequences of 1000 bases long contain exactly 2 mutant bases.

(c) What is the probability of having the 500th base and the 888th base mutated in the pool of DNA which has only two mutations?

(d) What is the probability of having two mutations side-by-side in the pool of DNA which has only two mutations?

So now going to the 3rd question, calculate how many combinations of DNA sequences of 1000 bases long contain exactly 2 mutant bases so it should be exactly 2 mutant bases, not more than that. So since it has to get exactly 2 mutant bases, so what was the earlier question when we said N equal to 2? That time it was out of 1000 bases, there are 2 mutations ok so how many possible ways we can get that, the question is how many possible ways we can get that. Calculate how many combinations of DNA sequences are there. So it is just asking not the probability, he is asking the combination so that is straightforward.

(Refer Slide Time: 25:36)



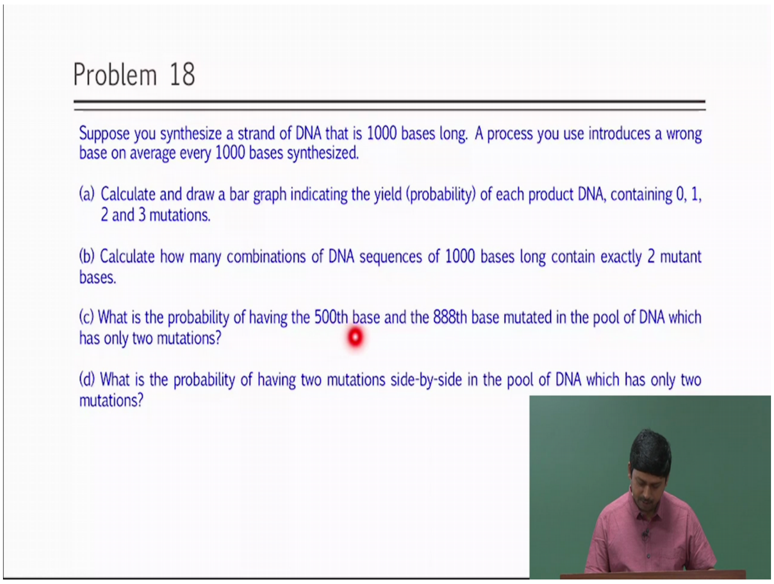
(b) $W = {}^{1000}C_2$

$= \frac{1000!}{998!\,2!} = \frac{\overset{500}{\cancel{1000}} \times 999}{\cancel{2}}$

$= 499500$

So it will be W is out of 1000 we are going to get 2 mutations which is 1000 C 2 so it is 1000 factorial by 998 factorial and 2 factorial which is 1000 into 999 divided by 2 is 500 into 999 is this.

(Refer Slide Time: 26:06)



Problem 18

Suppose you synthesize a strand of DNA that is 1000 bases long. A process you use introduces a wrong base on average every 1000 bases synthesized.

(a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2 and 3 mutations.

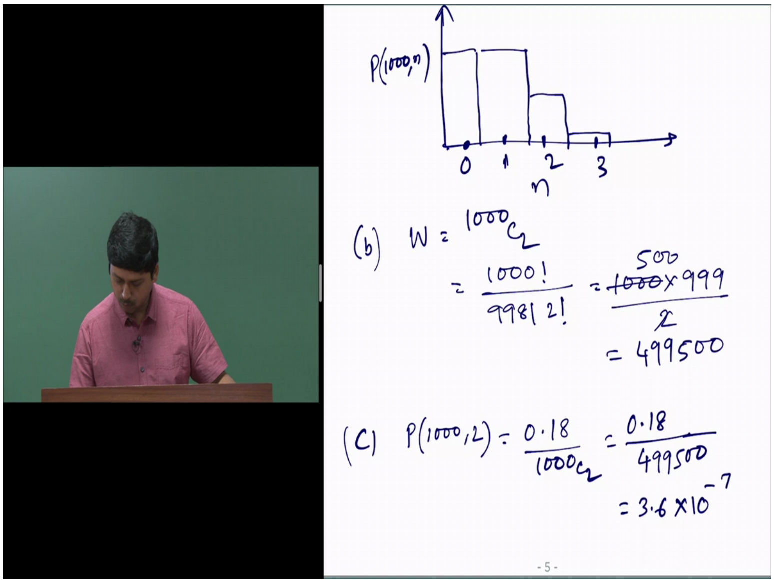
(b) Calculate how many combinations of DNA sequences of 1000 bases long contain exactly 2 mutant bases.

(c) What is the probability of having the 500th base and the 888th base mutated in the pool of DNA which has only two mutations?

(d) What is the probability of having two mutations side-by-side in the pool of DNA which has only two mutations?

Now question number 3, what is the probability of having 500th base and 88th base mutated in the pool of DNA which has only 2 mutations. So it is specifying the mutations exactly in the right place which means that any 2 mutations, we have seen the probability that there are any 2 mutations which means when P is 1000, 2 then it is approximately 0.18 but that time we considered mutations can be at any place, it can be at 1, 2, it can be at 3, 4, it can be at 3 and 27, it can be at like any 2 places, and how many such possible combinations are there? 1000 C 2, now those 1000 C 2 we have multiplied in order to get that, now we have to divide by those 1000 C 2 because here the positions of the mutations are very specific so they cannot interchange.

(Refer Slide Time: 27:51)



(b) $W = {}^{1000}C_2$

$= \dfrac{1000!}{998!\,2!} = \dfrac{\overset{500}{\cancel{1000}} \times 999}{\cancel{2}}$

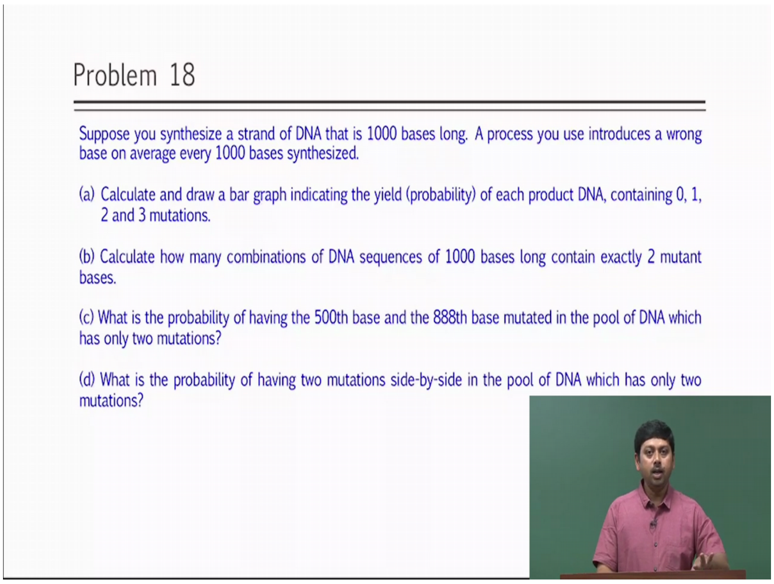
$= 499500$

(C) $P(1000,2) = \dfrac{0.18}{{}^{1000}C_2} = \dfrac{0.18}{499500}$

$= 3.6 \times 10^{-7}$

So now we are going to divide by that 1000 C 2 which is 0.18 into 1000, we have already done that of course, it will be 499500, let us see how much that comes, ok so it is 3.6 into 10 to the power – 7.

(Refer Slide Time: 28:03)



Problem 18

Suppose you synthesize a strand of DNA that is 1000 bases long. A process you use introduces a wrong base on average every 1000 bases synthesized.

(a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2 and 3 mutations.

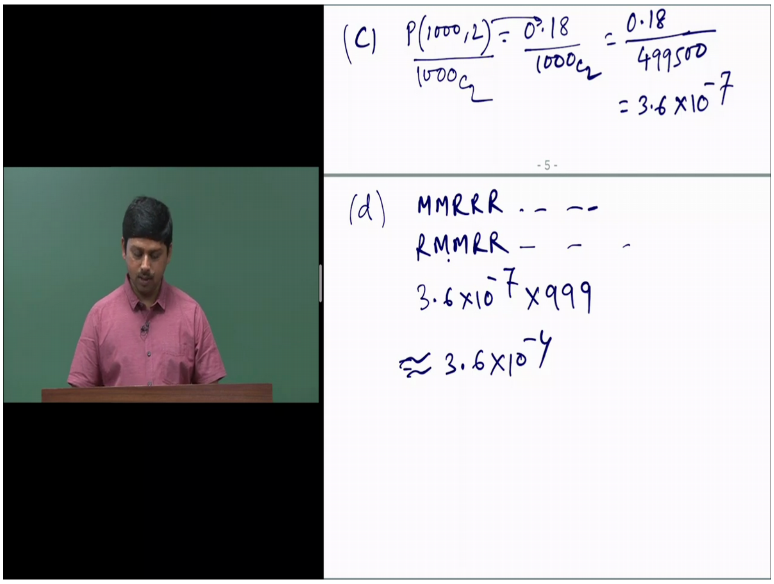
(b) Calculate how many combinations of DNA sequences of 1000 bases long contain exactly 2 mutant bases.

(c) What is the probability of having the 500th base and the 888th base mutated in the pool of DNA which has only two mutations?

(d) What is the probability of having two mutations side-by-side in the pool of DNA which has only two mutations?

So now we are going to go to the last part of this question, what is the probability of having to mutations side-by-side in the pool of DNA which has only 2 mutations. Now earlier it was very specific mutations, before that the position of mutations were not mentioned, now they are saying that they are together which means the mutations can be at 1 and 2 positions or 2 and 3 positions or 3 and 4 positions, so it can be any of those 999 positions because it cannot be in the 1000 position because there are only 1000 sequences so it cannot be occupying 1000 and 1001 place, but it can occupy 999th place and 1000th place, 1st place and 2nd place and things like that.

(Refer Slide Time: 28:37)



(c) $\frac{P(1000,2)}{{}^{1000}C_2} = \frac{0.18}{{}^{1000}C_2} = \frac{0.18}{499500}$

$= 3.6\times10^{-7}$

-5-

(d) MMRRR . . . . .

RMMRR . . . .

$3.6\times10^{-7}\times999$

$\approx 3.6\times10^{-4}$

So the mutations are together so it is like this, so mutation is let us say denoted by M and right bases are denoted by R so it can be RRR 998 of them or it can be R and then MMRRR and rest of them so always they are together, so that means there are 999 places that there can be there. So in the previous case we did not even specify any place, it was totally it is very specific and before that it could have been anywhere else, so taking the answer from C where no other possibilities were actually given, we can just multiply 999 with this answer and then we can get back so this is this answer is of course only for this because we are dividing by that so this is we should write 1000 C 2.

Now take this result from the earlier result which is 3.6 into 10 to the power – 7 where it is very specific mutation and then there are 999 more possibilities rather in this case and then that is what it is going to give us, so it is very approximately 1000 so let us say it is very-very approximately 3.6 into 10 to the power – 4 that will be the answer for this ok ok.