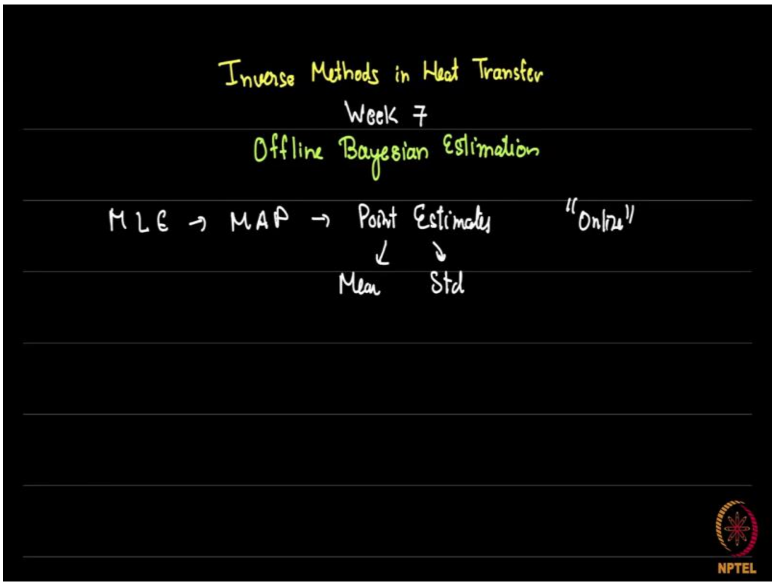
Inverse Methods in Heat Transfer

Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology - Madras

Lecture - 39
Offline Bayesian Estimation

(Refer Slide Time: 00:19)



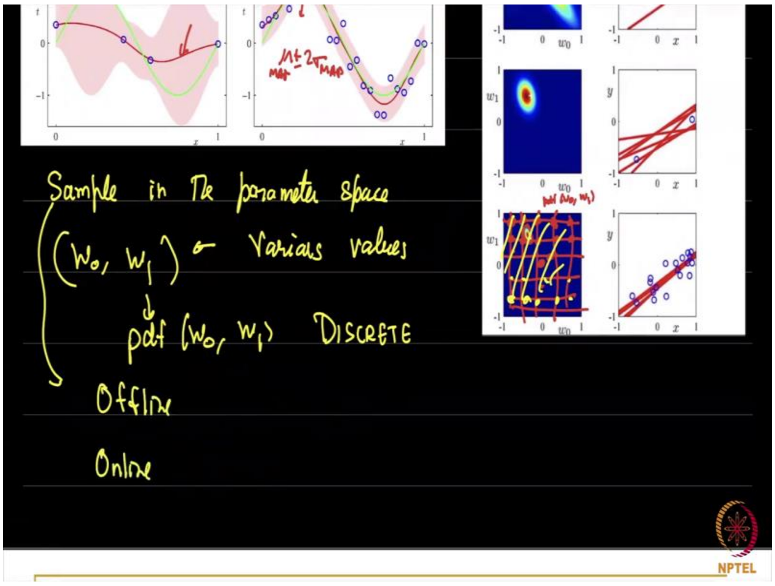
Inverse Methods in Heat Transfer
Week 7
Offline Bayesian Estimation

MLE → MAP → Point Estimates "Online"
Mean Std

Welcome back, we are in week seven of inverse methods in heat transfer. We talked about what Bayesian estimation meant in the last video and I am going to discuss about a really simple approach called an offline Bayesian estimation. In this we go through the same process as last time, which is we evaluate some likelihood, from there we it evaluates the posterior and based on that, we basically find out Point estimates for the mean and the standard deviation. This will become clearer when I talk about the method right now.

Now offline basic and estimation is having an opposite or a corresponding other thing called online Bayesian estimation. I will talk about both of these shortly in this video itself and we will take a simple example from the slab and then proceed from there.

(Refer Slide Time: 01:27)



Sample in The parameter space

(w_0, w_1) ← Various values

↓

pdf (w_0, w_1) DISCRETE

Offline

Online

So, just to recollect where we are going again this is a slide or a picture from the Christopher Bishop book, remember that let us say we have some data generation this is the ground truth. This green line here that you may or may not be able to see is a sinusoid and this is basically where we generate the data from and we add some noise to it. So, once we add some noise to it based on this, this is our estimate.

So, this is our truth or reality, this is noisy data and this is our model of what happens our inverse estimate model is. Now the band here the pinkish or peachish band that you see here is basically our Bayesian uncertainty. Given what we know we are assuming for example this is a ninth order polynomial let us say and given this data we really are uncertain even though we said this is the line the line could be anywhere in between this that is what Bayesian estimates tell us.

Similarly, if you give two points the uncertainty region becomes a little bit lower and as you add more and more points you can see that our uncertainty becomes smaller and smaller and finally you have a very small region here. Now what we want to do as we finish todays or this particular video is to give this red line as the mean and this up and down as mu plus minus two Sigma.

So, we want an estimate for this which is which will usually turn out to be the map estimate or maximum a-posterior estimate the maximum a posterior estimate of the mean map we also want to give the standard deviation. So, the variation which is also Sigma so, we want to give both these estimates and we will see that it is a simple trick we can do that rather than doing a full sort of Bayesian calculation which I showed you in the last video with the integration of the denominator etcetera.

Now when we do that, the method that we will be using which currently the method that we are using in this video will be what is called an offline Bayesian. The idea is very simple. Now notice this picture here. So, this is for a different problem all together, this was for the sinusoid explanation, this is for that simple linear function estimation. we know that finally we want to converge here this gives us w_0 & w_1 it is giving us the PDF of w_0 & w_1 this is the joint PDF of w_0 & w_1.

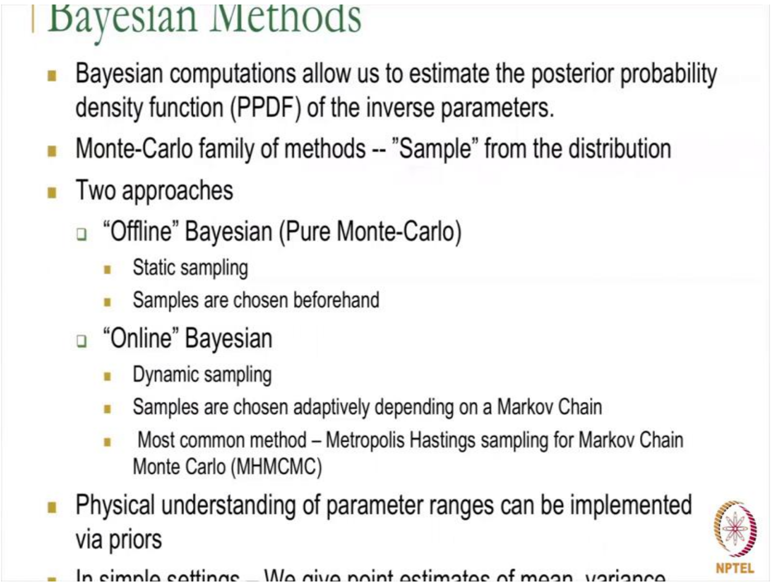
In order to get that we will actually have to sample what does it what do I mean by sample that what is the probability of let us say w_0 equal to 0 and w_1 equal to 0, you sample it here and give a value. You cannot give the full function because theoretically it is very hard to estimate as I told you we cannot get theoretical estimations other than Gaussian or stuff like that some easy other than those easy cases you actually have to calculate this computational.

Now the way we are going to do it in this video is we will essentially split it into a grid, much like if you have done a CFD course if it is in some sense like that. And at each point we will evaluate the likelihood, the posterior or the likelihood, the prior and the posterior and based on that we calculate what is known as the posterior PDF, PPDF. The idea is just like what I said before use the previous posterior as the new prior and keep on multiplying it that is one idea. we will use a simpler idea in fact as far as this video is concerned.

The idea therefore is to sample in the parameter space. Now remember where we are sampling in the parameter space that is find out various values of w_0 & w_1 and evaluate PDF at that specific w_0 & w_1 and overall if you cover a lot of values of w_0 & w_1 you will cover the entire space. So, instead of getting this continuous nice smooth distribution you get discrete distributions.

Now there are two ways of doing this sampling. That is what is known as the offline method and then there is known as the online method. So, let us come to what these two methods are.

(Refer Slide Time: 06:46)



Bayesian Methods

- Bayesian computations allow us to estimate the posterior probability density function (PPDF) of the inverse parameters.
- Monte-Carlo family of methods -- ”Sample” from the distribution
- Two approaches
  - ❑ “Offline” Bayesian (Pure Monte-Carlo)
    - Static sampling
    - Samples are chosen beforehand
  - ❑ “Online” Bayesian
    - Dynamic sampling
    - Samples are chosen adaptively depending on a Markov Chain
    - Most common method – Metropolis Hastings sampling for Markov Chain Monte Carlo (MHMCMC)
- Physical understanding of parameter ranges can be implemented via priors
- In simple settings – We give point estimates of mean, variance

So, as I have said just now, this Bayesian computation allows us to estimate the posterior probability density function which we are going to call PPDF, P for posterior probability density function for the inverse parameters which we are evaluating w_0 & w_1 for example. And this family of methods these family of computational methods are known as Monte Carlo methods these Monte Carlo methods have a simple idea. you sample from the distribution means you sample from the space of parameters.

And you calculate the value of the probability density function at that point. usually, more specifically, we calculate when we say distribution only the numerator, this is in some sense non-normalized. if you remember we kept on using the fact that the posterior probability is proportional to the prior probability multiplied by the likelihood usually there was a denominator. So, this is just of course the numerator, there is a denominator here, which is probability of x and we do not ever really calculate it.

So, Monte Carlo methods are clever at finding out the normalized version of this without ever actually explicitly calculating the denominator. Now why that is etcetera would make this into a full course on probability and we do not have the time for that. So, as of now I am just going to give you some sort of hand waving arguments and sort of simple Arguments for how we calculate this.

Now as I said there are two ways of this sampling the real trick here is how do you cover this entire space of parameters efficiently and quickly jump into the place, where the probability is high, that is the basic trick. Now the first trick is basically a simple trick it is not even a trick it is a static sampling. It just what you do is this? If you have a large space to cover, you first take big steps. So, let us say we split it into 16. and after you split it into 16 you find out that most of the highest probabilities in this region.

Then you zoom in and split that let us say into nine further let us say then you find out that this is where it is highest zoom in further split that into if you want 9 or 16 or however how much ever and so on and so forth that is basically how you isolate this probability density functions. So, that is the offline method. The offline method is a static sampling method what is meant by Static the samples are chosen beforehand initially.

For the very first sample, so, this initial 16 steps that we do we have this exact 16 points we know exactly that we are going to sample. In the online sampling method, which I will discuss later on this week, what we do is dynamic sampling. we first sample at one point and based on the value that comes at that point go to the next point this is somewhat like gradient descent. You first find out the value of a function and based on what that value is.

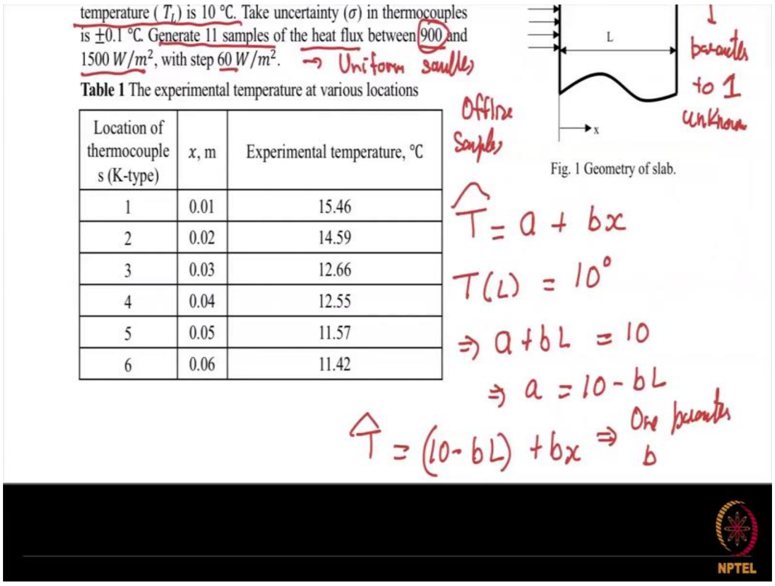
You could go to a next place where you are more likely to find a nice higher probability rather than see just imagine that you are in this space and most of the time you are sampling here only that is waste. If I spend 2000 samples here that is computational expense that is wasted, I can see that most of the probability is concentrated here if you can if I can somehow jump easily into this space and I can evaluate this quite fast then that would make a lot more sense rather than statically sampling in the entire space.

So, these are the two approaches, thus the offline approach which is what I am going to show in this video and the online approach which I will show later on this week. And the standard algorithm in the online approach is what is called MHMCMC which is Metropolis Hastings Markov Chain Monte Carlo. Now Monte Carlo is historically it is just a casino and a lot of gambling was done it is said that the father of the Monte Carlo methods.

Which is kind of the online Bayesian method that I am showing simple static sampling was somebody called Stanislav Ulam. He came up with this method during while these days were making the atom bomb during 1940s and Stanislav ulam's father lost a lot of money in gambling which is and he sort of made up this method and called it Monte Carlo because it is supposed to evaluate probability distributions in an efficient fashion.

Now as we saw in the previous videos the physical understanding of what the parameter ranges, we want our parameters to wheel can be implemented by our priors we are going to give in a very simple setting we are going to give Point estimates of the mean and the variance.

(Refer Slide Time: 12:22)



temperature (T_L) is 10 °C. Take uncertainty (σ) in thermocouples is ±0.1 °C. Generate 11 samples of the heat flux between 900 and 1500 W/m^2, with step 60 W/m^2. → Uniform samples

Offline Samples

Table 1 The experimental temperature at various locations

Location of thermocouples (K-type)	x, m	Experimental temperature, °C
1	0.01	15.46
2	0.02	14.59
3	0.03	12.66
4	0.04	12.55
5	0.05	11.57
6	0.06	11.42

Fig. 1 Geometry of slab.

$$\hat{T} = a + bx$$

$$T(L) = 10^\circ$$

$$\Rightarrow a + bL = 10$$

$$\Rightarrow a = 10 - bL$$

$$\hat{T} = (10 - bL) + bx \Rightarrow \overset{\text{One parameter}}{b}$$

NPTEL

And we are going to assume the following three Gaussians that the prior data or prior estimate of parameters is a Gaussian. We are going to assume that the data has Gaussian noise and we are going to assume that the posterior or the Bayesian estimate finalling is also Gaussian. So, all three are assumed to be Gaussians. So, let us come to our favourite problem once again the simple slab problem I have modified it slightly here in order to make it computationally feasible using probabilistic methods.

So, the problem in the first instance will seem exactly the same as before I will explain how I have modified it slightly. Once again, a steady state heat conduction problem you are asked to estimate the heat flux q. But I am asking you in some more detail, I am not asking you to just find out who I am asking you to find out mean q. I am also asking you to give you give me a map - maximum a posteriority and I am also asking you to find the standard deviation given that there is no prior.

This is somewhat of a mild tricky question that is we have no idea about the heat flux from before, but I am asking you to find out the mean the MAP and the standard deviation in the heat flux. The experimental measurements are exactly the same as before. But the two modifications we have made are this we are giving you the boundary condition at this area and I have told you exactly that this temperature is 10 degrees Centigrade.

Now I can also give it as 10 degrees Centigrade with some uncertainty that that is not given. you are given that this is exactly 10 degrees Centigrade why did we do that. Once I give you one temperature and I will show you what happens, this reduces the parameter to just one unknown. Now remember all throughout when we were doing linear regression, we had two unknown parameters in this case we have one single unknown parameter.

And why did I reduce it to one as you saw in the plot here, I have to search for the parameter in a two-dimensional space now that is a little bit painful and it is hard for you to visualize this probability density function also. So, just in order to simplify it for classroom purposes, we have made the parameter space to only one I will show you how it becomes only one shortly. Also, we have also told you the uncertainty in the thermocouples.

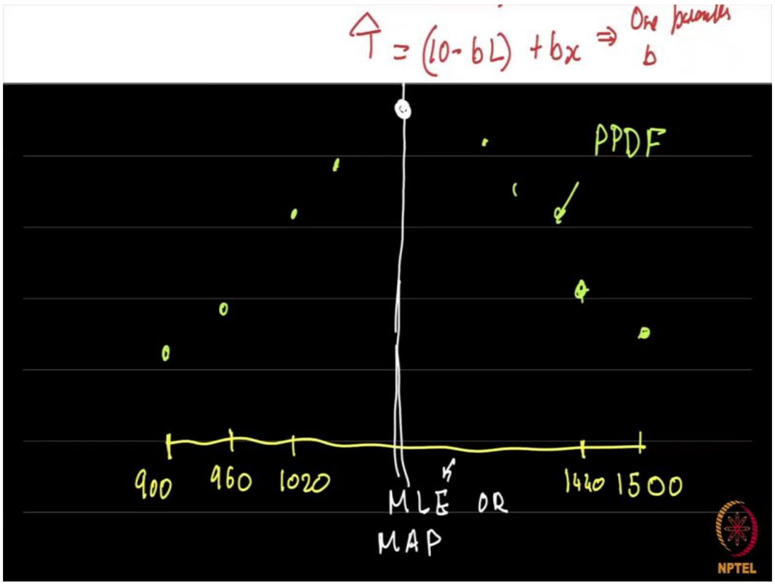
So, each one is an uncertainty a standard deviation of plus minus one degree Centigrade. Now you are asked to take samples. you will now see you are asked to generate level samples of the heat flux. And now notice I have actually given you a range of heat flux this itself in some cells gives you some idea of where you are searching physically, but this is a uniform sample if you take 60 of these, you'll get a range of 600.

So, you get 11 samples from 900 to 1500. these are uniform samples; these are also offline samples that is I did not tell you start with a sample of 900 and then go as you please which we will do when we come to the Metropolis testings algorithm. We will start with just one initial guess, just like we did with linear regression you start with an initial guess and then or with non-linear regression also.

And then you start finding your way out that is more like online sampling or the Marco chain Monte Carlo algorithm. So, let us try and solve this problem sequentially. So, first remember our model $\hat{T}$ or what we called $\hat{y}$ was $a + bx$ or $w_0 + w_1 x$. But we have one additional piece of information we know that $T(L) = 10℃$.

So, this gives us that $a + bL = 10$ which means you can say that $a = 10 - bL$. So, you might be able to say $\hat{T} = (10 - bL) + bx$. So, there is just one parameter b for this problem. So, this is a single parameter problem which we have reduced our problem. Now what are we going to do.

(Refer Slide Time: 18:09)

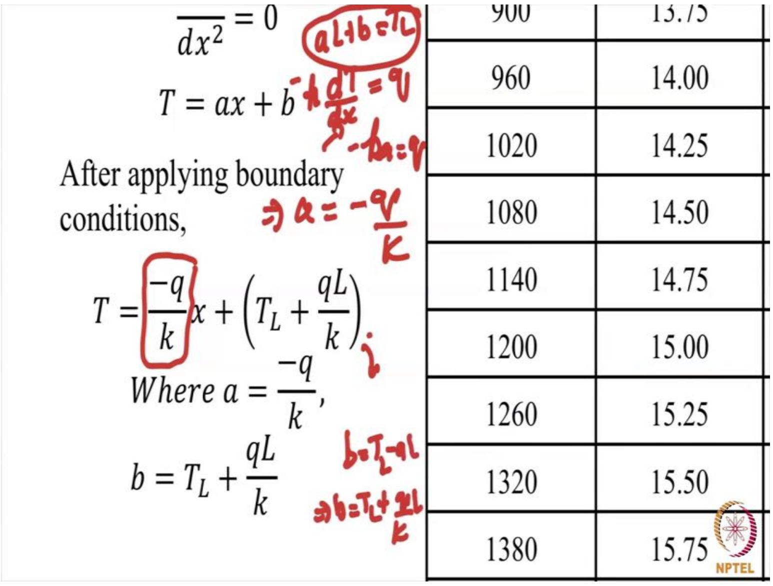


So, we are going to search in the space 900 to 1500 and we are going to keep 11 samples in between. So, let us say 960, 1020 so on and so forth until 1440. Now ideally what you want to do is to build this entire posterior PDF, what you want to build is this entire thing however you will not be able to do that. you will only be able to sample and get 10 points or 11 points. So, something of this sort is all you will be able to manage.

Now once we get here what do we do, we simply find out where it is maximum. Let us say it is maximum here then I will say this is my maximum likelihood estimate, in case I have Incorporated the prior or it is the map sorry in case I have not Incorporated the prior then it is MLE, if I incorporated the prior then it is mapped all right. Now for each of these points generated. So, let us see how each one of these will be generated now.

So, what you have to do is as follows, let me show you the entire model here, so let us concentrate here.

(Refer Slide Time: 19:40)



$$\frac{}{dx^2} = 0 \quad \boxed{aL+b=T_L}$$

$$T = ax + b \quad -k\frac{dT}{dx} = q$$

$$\Rightarrow -ka = q$$

After applying boundary conditions,

$$\Rightarrow a = -\frac{q}{K}$$

$$T = \boxed{\frac{-q}{k}}x + \left(T_L + \frac{qL}{k}\right); $$

$$Where\ a = \frac{-q}{k},$$

$$b = T_L + \frac{qL}{k} \quad b = T_L - aL$$

$$\Rightarrow b = T_L + \frac{qL}{K}$$

900	13.75
960	14.00
1020	14.25
1080	14.50
1140	14.75
1200	15.00
1260	15.25
1320	15.50
1380	15.75

So, sorry I wrote it as $a + bx$ here, it is written as $ax + b$. I will just repeat the calculation once more given. So, remember this is our model or forward model which is entirely from physics. So, physics says $\frac{d^2T}{dx^2} = 0$ therefore $\hat{T} = ax + b$, after you apply the boundary conditions just like I did earlier the boundary conditions say that $aL + b = T_L$. So, once you apply that and plug that back in you will get this value here.

So, what you will get is both in terms of q. So, you get both these terms of q, how do you get this in terms of q you say,

$$-k\ \frac{dT}{dx} = q$$

which means,

$$-ka = q$$

So, this means a equal to,

$$a = \frac{-q}{k}$$

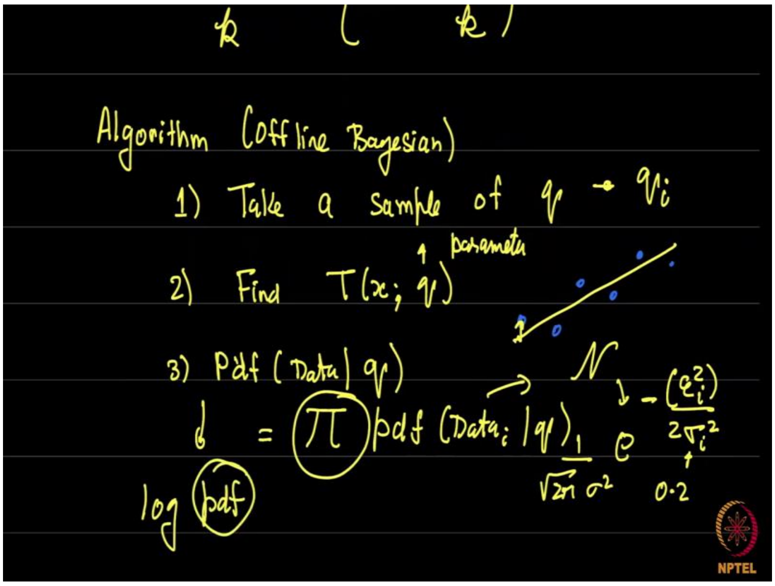
and then b you can get from here, how do you get this you say,

$$b = T_L - aL$$

$$b = T_L + \frac{qL}{k}$$

So, you get this all right.

(Refer Slide Time: 21:22)



Algorithm (Offline Bayesian)

1) Take a sample of $q \rightarrow q_i$

2) Find $T(x;\ q)$

3) $\text{Pdf}(\text{Data} \mid q)$

$$= \prod \text{pdf}(\text{Data}_i \mid q)_i \quad \frac{1}{\sqrt{2\pi}\,\sigma^2}\, e^{-\frac{(\varepsilon_i^2)}{2\sigma_i^2}}$$

log (pdf)

So, now you have a new expression for T that is let me go back here I can write this as,

$$T = \frac{-q}{k}x + (T_L + \frac{qL}{k})$$

So, what's the algorithm? The algorithm for offline Bayesian this will become a little bit clearer once I show you the code in the next video but the algorithm for offline Bayesian is very easy. you take a sample of q. So, let us call this q_i. So, let us say I start here I take a sample of q which is 900.

Now, based on this q find T; so, T corresponding to this q, T of x which depends on this q. Now q is the parameter. So, you will be based on this q you will draw a line. Now the real data is somewhere around this line. So, you had six data points around this line you can. Now evaluate the PDF of the data given this value of q. How do you evaluate this, we already saw this. This is the product of original each point even this q and each of this was a normal.

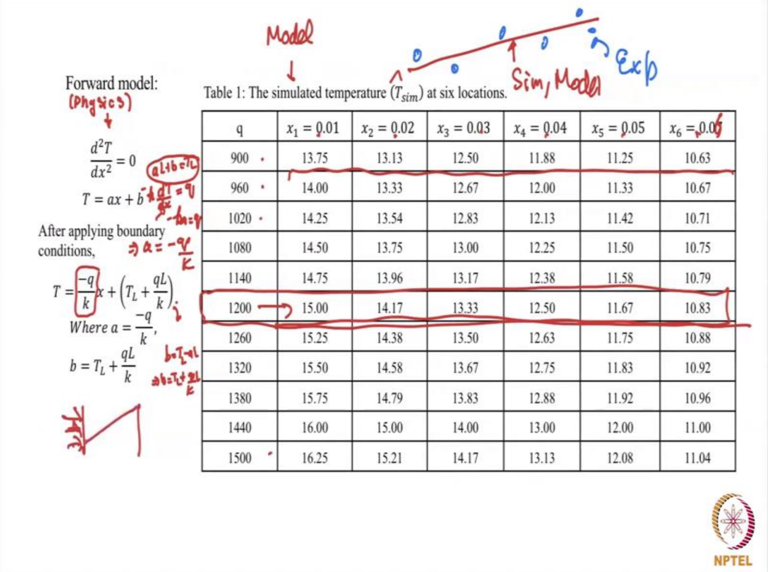
So, let us see this. So, remember this normal, this normal was e to the power minus the gap between my prediction and the original. So, this was,

$$e^{\frac{-\varepsilon_i^2}{2\sigma_i^2}}$$

The sigma was already given as the uncertainty in the thermocouple which is fine. So, we can find this and if you wish we can multiply this by 2π. Now in what follows I am going to do this really speaking what you should evaluate is the log of the PDF.

And I will tell you why once I show these numbers in the upcoming slides. So, let us just do it naively and go over point by point.

(Refer Slide Time: 23:58)



Forward model:

$$\frac{d^2T}{dx^2} = 0$$

$$T = ax + b$$

After applying boundary conditions,

$$T = \boxed{\frac{-q}{k}}x + \left(T_L + \frac{qL}{k}\right),$$

$$Where\ a = \frac{-q}{k},$$

$$b = T_L + \frac{qL}{k}$$

Table 1: The simulated temperature (T_{sim}) at six locations.

q	$x_1 = 0.01$	$x_2 = 0.02$	$x_3 = 0.03$	$x_4 = 0.04$	$x_5 = 0.05$	$x_6 = 0.07$
900	13.75	13.13	12.50	11.88	11.25	10.63
960	14.00	13.33	12.67	12.00	11.33	10.67
1020	14.25	13.54	12.83	12.13	11.42	10.71
1080	14.50	13.75	13.00	12.25	11.50	10.75
1140	14.75	13.96	13.17	12.38	11.58	10.79
1200	15.00	14.17	13.33	12.50	11.67	10.83
1260	15.25	14.38	13.50	12.63	11.75	10.88
1320	15.50	14.58	13.67	12.75	11.83	10.92
1380	15.75	14.79	13.83	12.88	11.92	10.96
1440	16.00	15.00	14.00	13.00	12.00	11.00
1500	16.25	15.21	14.17	13.13	12.08	11.04

So, let us say that I have this range of q values going from 900 to 1500, you can see there are one two three up till 11 points and I have my data set, I have one, two, three, four, five, six, six points that I have this, this should be 0.06 not 0.07. Now at these points I calculate my predicted temperature. So, simulator temperature is the model temperature depending on this expression. Now, what this means is notice for each parameter I have to calculate at each point.

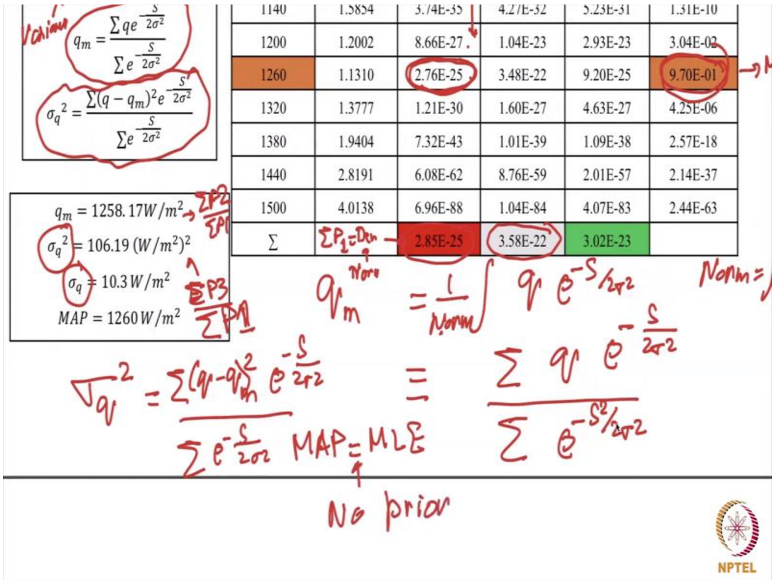
Now imagine if I had two parameters this will be very expensive. So, if I have two parameters then there will be a large grid. So, if I take let us say my right-hand boundary condition was also not known then I have a and b is two unknowns, I take 10 values of a, 10 values of B which means I will have 100 by 100 grid and at each value I would have to generate six temperature points.

What does this correspond to? So, please understand this when q was 900, it will give a line and these six are the values on that line and as we will see we are going to compare that with ground truth. This is experiment this is I am calling it simulation or the model which we used to call $\hat{T}$ but I am going to call it T Cell right. Now just to remember that this is a simulation that we are doing with what is called Monte Carlo.

I just simply sampled this queue from various places. Now each one of these gives a different temperature distribution this gives one temperature distribution when I choose it to be let us say 1200 the way to read, it is to start from the value of q say that if the heat flux was so much, what would the temperature look like. So, that is basically what we are doing effectively we are doing a hit and trial method.

We say if I know that the heat flux is between 900 and 1500 for each one of them how probable is it that the heat flux was 1200 or 900 or 960. So, the first step was to calculate the simulated temperature. The next step I have written just this let me just finish the algorithm find the max of all q_i. So, find out the maximum PDF value and that would be the MLE in case we have not incorporated the prior it could be the map in case we have Incorporated the prime.

(Refer Slide Time: 26:53)



Variance

$$q_m = \frac{\sum q e^{-\frac{S}{2\sigma^2}}}{\sum e^{-\frac{S}{2\sigma^2}}}$$

$$\sigma_q{}^2 = \frac{\sum (q - q_m)^2 e^{-\frac{S}{2\sigma^2}}}{\sum e^{-\frac{S}{2\sigma^2}}}$$

1140	1.5854	3.74E-35	4.27E-32	5.23E-31	1.31E-10
1200	1.2002	8.66E-27	1.04E-23	2.93E-23	3.04E-02
1260	1.1310	2.76E-25	3.48E-22	9.20E-25	9.70E-01
1320	1.3777	1.21E-30	1.60E-27	4.63E-27	4.25E-06
1380	1.9404	7.32E-43	1.01E-39	1.09E-38	2.57E-18
1440	2.8191	6.08E-62	8.76E-59	2.01E-57	2.14E-37
1500	4.0138	6.96E-88	1.04E-84	4.07E-83	2.44E-63
$\sum$	$\sum P_1 = \text{Den}$	2.85E-25	3.58E-22	3.02E-23	

$q_m = 1258.17 W/m^2 \rightarrow \frac{\sum P2}{\sum P1}$

$\sigma_q{}^2 = 106.19\ (W/m^2)^2$

$\sigma_q = 10.3\ W/m^2$ $\frac{\sum P3}{\sum P1}$

$MAP = 1260\ W/m^2$

$$q_m^{\text{Norm}} = \frac{1}{\text{Norm}} \int q e^{-S/2\sigma^2} \qquad \text{Norm} = \int$$

$$\sigma_q^2 = \frac{\sum (q - q_m)^2 e^{-\frac{S}{2\sigma^2}}}{\sum e^{-\frac{S}{2\sigma^2}}} \equiv \frac{\sum q e^{-\frac{S}{2\sigma^2}}}{\sum e^{-S^2/2\sigma^2}}$$

MAP = MLE

No prior

So, let us come to this, this is the simulation. Now this simulation value obviously is not going to match the actual experimental value. So, now notice if q was 900, add these six locations, let me change this once again to in zero six, at these six locations you will see some values, these differences are high here they become lower progressively, but then they might become higher somewhere else, you will notice just by eyeballing it.

That somewhere around either here or here is where you are getting the lowest temperature differences in absolute value. But we are not happy with that the actually evaluate. So, this of course is the equivalent of my ε_i. ε_i is whatever being our model value divided by the actual value you can do plus or minus of this since we are going to square it, it does not matter. What we do now here is where the majority of the calculation is.

So, it might look a little bit confusing, but let us go step by step. So, you remember this back when we did linear regression, we had this value this is of course the sum of squared errors.

We can call it ε or E also. I will keep ourselves to the original notation we used which was S without the average the average was called J the loss function, but just the sum of the errors is whatever was our model value minus whatever was the True Value you square that and you sum it and that you get as S.

And if you remember,

$$pdf \propto e^{\frac{-S^2}{2\sigma^2}}$$

So, that is what sits here this is what we call P_1. P_1 is just this numerator why only the numerator, because finally we really do not care about the denominator in this process that will finally come out from the normalization. So, we are ignoring notice here we are ignoring the $\frac{1}{\sqrt{2\pi\sigma^2}}$ you can put Sigma outside of the group that we are ignoring.

This minus might not be clear there is actually a minus sitting here which you can see if you zoom in it. So, there is a minus there. So, please notice that. So, I will just write it a little bit more explicitly, there is a minus there. we are also going to calculate a quantity called P_2 Now why is P_2 being calculated because ultimately P_2 is the mean which is calculated as whatever quantity we want.

Suppose we want the quantity which is the temperature or in this case the heat flux. So, this is $\int \frac{qe^{\frac{-s^2}{2\sigma^2}}}{Norm}$. this is not P_2 but this is E(q) equal what is expected value of q? This is the mean value of q that we are going to calculate. Remember our definition of expectation as well as variance.

So, we are going to look at not only is what is q? What is the mean value of q. Now how do how does it relate to the pictures that we had seen earlier. So, if you come here the mean or the MAP. Now depending on how we do it you might ask either end up at the mean or at the MAP, but something like the middle value around which everything else clusters for everything else varies is what we are calculating.

Just like for any probability distribution we are actually calculating the mean value via a simple. So, this is the trick that we use for a Bayesian estimate. we directly calculate the mean rather than calculating the full Bayesian calculation. So, we calculate the maximum and let me add find the mean of the parameter. So, let us say expected value of q,

$$Var[q] = E(q) = \int \frac{qe^{\frac{-s^2}{2\sigma^2}}}{Norm}$$

and we will forget the normalization.

So, that we will do right at the end and we will also find the variance of q if you remember once you find out q. So, this is if we call this mu, then this is,

$$Var[q] = \int (q - q_m)^2 e^{\frac{-S^2}{2\sigma^2}}$$

you do this over the entire range. So, let us come back to this here. This calculation can actually be done at the end after identifying some candidate heat fluxes also, but let us do it in a systematic fashion.

So, this P_2 here is used to calculate the mean. So, P_1 is used to calculate the maximum, P_2 is calculated to use to calculate the mean and P_3 is used to calculate the variance. Now notice how we are doing the normalization. qm is actually speaking the normalized integral of,

$$q_m = \frac{1}{Norm} \int qe^{\frac{-S^2}{2\sigma^2}}$$

But instead of doing that the computational equivalent and the Bayesian equivalent is,

$$\equiv \frac{\sum qe^{\frac{-S^2}{2\sigma^2}}}{\sum e^{\frac{-S^2}{2\sigma^2}}}$$

Because what comes as the normalization is the,

$$Norm = \int e^{\frac{-S^2}{2\sigma^2}}$$

so, instead of doing that we do this. Similarly, when we want to find out the variance the variance is calculated as,

$$\sigma_q{}^2 = \frac{\int (q - q_m)^2 \, e^{\frac{-S}{2\sigma^2}}}{\int e^{\frac{-S}{2\sigma^2}}}$$

I think I have been writing s Square by mistake which should just be s, s already has the square setting down.

So, please correct your notes if you have been doing that minus s divided by integral of minus s by 2 Sigma Square. Now in practice, these integrals will be replaced by sums as you can see.

Again, I will show you the code in detail in the next video but let us now look at how to do this with the data set that we have all right. Now before we go further notice the numbers that are coming out.

So, you can see very small numbers. So, in practice it actually makes sense to work with log PDFs, you know take the log back and then do the normalization but I am going to work with these small numbers just now just for convenience. So, so let us start with these small numbers. So, now notice what is happening, then I guess the heat flux of 900, the error this is the sum square error between my model and my ground truth was 6.28.

So, you can see this variation it starts decreasing till you reach this point and its minimum as at one two six zero and then it starts increasing again. So, you can already see that this is actually the best possible value, but we will talk about this also from the perspective of a posterior PDF also. So, you have this 1.1310. Now P1, what was P1? it was simply the PDF or the numerator of the PDF.

So, at this point it was some 10 powers minus 25, when you normalize it how do you normalize it, we will come to that. So, you see this P1 this Sigma P1 that you see at the bottom this is the denominator, this is the normalization constant why e to the power, Sigma of e to the power minus s by 2 Sigma Square sits here and this is the sum exactly are tend to do. So, this is the minimum that comes here. Now next is P2, P2 is used to calculate the mean. So, how are we calculating the mean the mean is being calculated as each q so 900 times P1.

So, that I sum over these values. So, 900 multiplied by e to the power minus s by 2 Sigma Square divided by e to the power minus s by to Sigma Square. So, you do this multiplication each time you get P2. this is the numerator of the P2 sum all those up you get 3.58 into 10 powers minus 22. So, divide 1 by the other you actually get mean of q. So, that is the way to actually calculate the estimate of the mean.

So, doing these things this way can be a little bit confusing. So, let me just show what P1, P2, P3 do and then we will look at the numbers just so, that if this becomes clear. So, for example P1 in the plot that I had drawn earlier let us draw it here this varies this is the probability distribution function or the numerator of the probability distribution function going from 900 to 1500.

So, what you should see when you plot P1 for each value of q notice on the x axis is q on the y axis is the PDF of q. So, that is what we are plotting how probable is this value of q that is what we have brought. So, you can see it is almost zero and Peaks at around is something of this sort. So, at around 1260 is this maximum value. So, you get the maximum value the normalized value at this point you can see these are the normalized value.

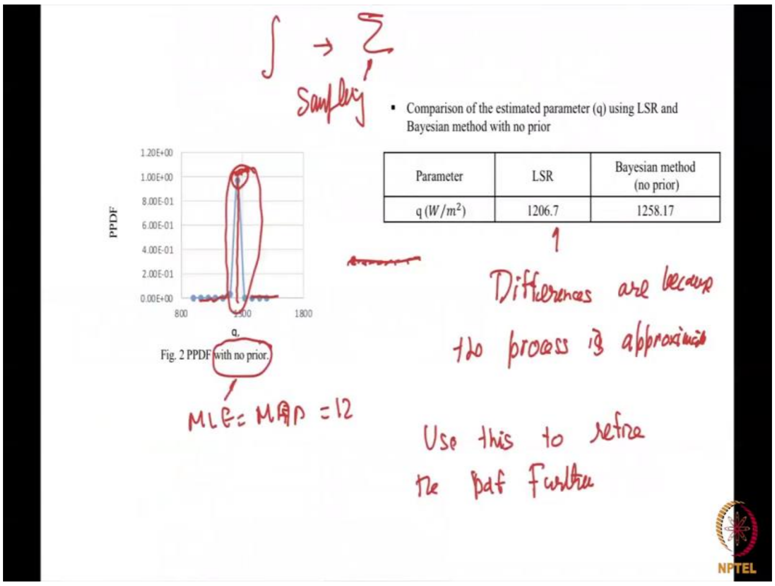
So, it starts at 10 powers minus 112 here really close to 0 and it comes to around 0.97 at this point and then rapidly drops off again to zero three. So, this PPDF you get by getting P1 divided by Sigma P1. So, that is how you get this. So, each of these values so, for example 2.76 into 10 powers minus 25 divided by 2.85 into 10 powers minus 25, will get you around 0.97. So, that is how you get the PPDF.

So, you know now how to calculate this and it is the PPDF that is plotted here. Now what about this value P2 what are we doing there. what we are calculating is contributions to the mean. once again remember the expectation of q is for each value of q you multiply by whatever error you got there and then normalize it and that is how you get the mean of q. So, what we are doing here in this Excel sheet this is actually clearer when you do this in MATLAB and hopefully you will see this a little bit more clearly when we see the next video but we weight this by the value of q.

So, this multiplied by 900, this multiplied by 960, this multiplied by 1020. So, we are saying that 900 occurred 10 powers minus 137 times, 960 occurred 10 power minus 101 times therefore in a weighted sense this is what occurred I take a sum of that and the sum of this is the numerator. So, some of this is the numerator here sum of q times e to the power minus s by 2 Sigma Square. So, this is the numerator the denominator is this.

So, the denominator is simply Sigma of P1. So, Sigma of P2 divided by Sigma of P1 gives me the mean value. Similarly, Sigma of P3 divided by Sigma of P1 gives me the variance. So, this value here if you see came from Sigma P2 divided by Sigma P1, this one came from Sigma P3 divided by Sigma P1. So, once we have that once we have Sigma q Square remember this is the variance you can actually get the standard deviation map and MLE are the same because there was no prior. So, here is what we see if we put stuff together.

(Refer Slide Time: 42:34)



$\int \rightarrow \Sigma$

Sampling

- Comparison of the estimated parameter (q) using LSR and Bayesian method with no prior

Parameter	LSR	Bayesian method (no prior)
q (W/m^2)	1206.7	1258.17

Fig. 2 PPDF with no prior.

MLE = MAP = 12

Differences are because the process is approximate

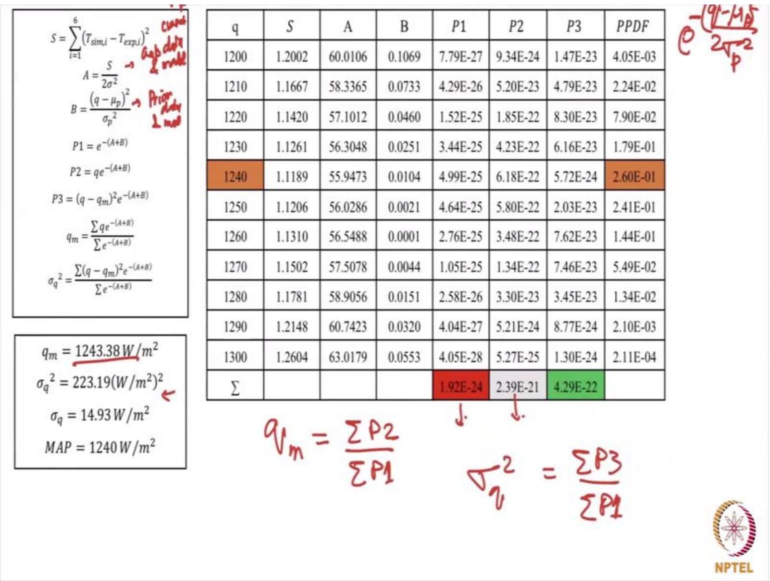
Use this to refine the pdf further

As I showed you there was no prior. This means MLE and map are the same if you simply looked at where the probability maximized, this place was approximately 1258 or somewhere around them so, 1260. So, this was 1260 amongst the points that we took it maximized at 1260. if you took a mean which came from that P 1 cap quantity by evaluating the integral all these are approximate integral it came to 1258.

And if you do a proper least square regression you will get somewhere around 1260. So, the difference is are because the process is approximate. what do I mean by the process is approximate we had to calculate integrals but instead we call calculated summations and these summations depend upon our sampling and we took a very simple sampling we went from 900 to 1500 in steps of 60 or so.

If we made the sampling a little bit smaller like we took you know smaller and smaller and smaller steps then this might look slightly better you will get closer and closer to least square regression. So, what you have now seen is a simple example of offline Bayesian sampling. we can refine this further um by going deeper into the regions where there are high probabilities.

(Refer Slide Time: 44:32)



$$S = \sum_{i=1}^{6}\left(T_{sim,i} - T_{exp,i}\right)^2$$

$$A = \frac{S}{2\sigma^2} \rightarrow \text{exp data \& model}$$

$$B = \frac{\left(q-\mu_p\right)^2}{\sigma_p^2} \rightarrow \text{Prior data \& mod}$$

$$P1 = e^{-(A+B)}$$

$$P2 = qe^{-(A+B)}$$

$$P3 = (q-q_m)^2 e^{-(A+B)}$$

$$q_m = \frac{\sum qe^{-(A+B)}}{\sum e^{-(A+B)}}$$

$${\sigma_q}^2 = \frac{\sum (q-q_m)^2 e^{-(A+B)}}{\sum e^{-(A+B)}}$$

$$\underline{q_m = 1243.38\,W/m^2}$$

$${\sigma_q}^2 = 223.19(W/m^2)^2 \leftarrow$$

$$\sigma_q = 14.93\,W/m^2$$

$$MAP = 1240\,W/m^2$$

q	S	A	B	P1	P2	P3	PPDF
1200	1.2002	60.0106	0.1069	7.79E-27	9.34E-24	1.47E-23	4.05E-03
1210	1.1667	58.3365	0.0733	4.29E-26	5.20E-23	4.79E-23	2.24E-02
1220	1.1420	57.1012	0.0460	1.52E-25	1.85E-22	8.30E-23	7.90E-02
1230	1.1261	56.3048	0.0251	3.44E-25	4.23E-22	6.16E-23	1.79E-01
1240	1.1189	55.9473	0.0104	4.99E-25	6.18E-22	5.72E-24	2.60E-01
1250	1.1206	56.0286	0.0021	4.64E-25	5.80E-22	2.03E-23	2.41E-01
1260	1.1310	56.5488	0.0001	2.76E-25	3.48E-22	7.62E-23	1.44E-01
1270	1.1502	57.5078	0.0044	1.05E-25	1.34E-22	7.46E-23	5.49E-02
1280	1.1781	58.9056	0.0151	2.58E-26	3.30E-23	3.45E-23	1.34E-02
1290	1.2148	60.7423	0.0320	4.04E-27	5.21E-24	8.77E-24	2.10E-03
1300	1.2604	63.0179	0.0553	4.05E-28	5.27E-25	1.30E-24	2.11E-04
Σ				1.92E-24	2.39E-21	4.29E-22	

$$e^{-\frac{(q-\mu_p)^2}{2\sigma_p^2}}$$

$$q_m = \frac{\Sigma P2}{\Sigma P1}$$

$$\sigma_q^2 = \frac{\Sigma P3}{\Sigma P1}$$

Now you can I am going to show you a very quick way of doing uh sort of incorporating the previous result. So, we took a course grid we got some result and we can use this to refine the PDF further. In fact, if you look at this PDF it is literally like all zero and just this one Peak. we would like to know a little bit more here in in case you want a slightly better estimate. So, the what the what we are going to do is this was our maximum likelihood estimate for the mean or not the maximum likelihood, this was just our mean estimate.

So, this was our Bayesian estimate for mean. we can use this to refine this region further by use using this as prior just like I showed you in the previous video. you use this as prior for the next calculation. Next, we will find it now an ultimate method of doing this is what is known as the online Bayesian, but let us just look at a casual way of doing this. Again, the same problem except we are asked to generate 11 samples between 1200 and 1300.

So, now we have decided by looking at this that most of the stuff is happening somewhere here. So, I will say instead of going from 900 to 1500 I will now go from 1200 to 1300 and I will take small step sizes. Now I can go 10 by 10 by 10. Now not only that in the previous case I had no prior, but I am going to use now a normal prior remember I am going to use a Gaussian prior normal simply means a Gaussian prior.

And in this case the Gaussian prior is going to have as new from the previous calculation. Now there are various ways of taking sigma, you could have taken the sigma that we had from the previous computation or if you have no information which is we are going to pretend we have no information if you have nothing, we will simply take this is just a rule of thumb take 10 percent of the previous mu.

So, we will just use the previous mu and pretend as if that Sigma is going to be the sigma of this process of the prior process. So, this is our first example of actually incorporating a base theorem in order to take a prior and multiply it with the previous PDF and create a posterior. So, the entire process looks almost exactly the same. So, the process is the same, first step is generating samples and from samples generate predictions model predictions.

Now you can imagine this can actually be done to a large-scale problem like weather problems. So, you assume that the cyclone location for example if the parameter you are solving for is let us say where is the location of the cyclone you assume a cyclone location and that will give you certain predictions for temperatures pressures and then you compare it with what you are measuring from the radar.

And then so, you assume something assume this value based on that regenerate the entire data set, then compare step two is compare with actual measurements. The entire inverse process that I told you make a guess compare it with ground truth we are literally doing that now. So, you again compare with your locations, the six locations where we made the measurements. So, these are those six locations again I should turn the skip 0.06.

And after that is where you generate PDFs. This is the third step; this is where most of the calculation is involved. Now here we have two steps of this calculation. what are the two steps the PDF is made up of the likelihood x given w and also the prior. The likelihood is simply e^{-A}, where A is given here,

$$A = \frac{S}{2\sigma^2}$$

This we already saw. But the posterior is e^{-B} where B remember is, or e to the power minus B this has to be a normal with mean whatever is the prior distribution mean and sigma of the prior.

So, this means B has to be this will look like $e^{\frac{-(q-\mu_p)^2}{2\sigma_p{}^2}}$, therefore B is going to be,

$$B = \frac{(q - \mu_p)^2}{2{\sigma_p}^2}$$

What does this physically mean A simply measures how well have you fit the data. So, the q you chose will give you an error which is sitting in s which will match the data but what this measure is how far is the q you chose from my previous assumption about how good q is.

So, mu p is actually a guess for what a q is based on what all you have seen in such slabs before or in our case based on what all we saw with the PDF that we saw before. So, please understand this delicate balance between data and prior. Data is what we are observing right. Now and prior is what we observed before. So, it makes sense that when you want to make a judgment right now, you should depend both on what you are seeing right now it is not all reality you know people before you have seen something you incorporate that too and that is how we make the software.

So, that is the power of the probabilistic approach. So, come back here. So, a is s by 2 Sigma square and b is q minus mu P Square. So, this is gap between data and model or gap between current data and model, this is gap between prior data and model. So, I hope that will you know make things a little bit more intuitive. So, now you see P1 is just like what we did before last time P1 was just minus a because we had no prior at all.

So, now P1 is equal to the E power minus a plus b, similarly P2 last time was just q into e to the power minus a, P is now going to be q into e to the power minus a plus b, p3 last time was q minus q m Square multiplied by E power minus a, this is q into e q minus q1 square e to the power minus a plus b everything else is exactly the same as before really nothing has changed except a wherever you had a.

Now, you have a plus b in other words wherever you had current data you. Now say current data plus sort of a compression of Prior data is what is sitting here all that together is sitting in this term d. So, we do the calculation and. Now you see that the maximum shifts a little bit because the prior shifts the maximum a little bit lower, the prior shifts the maximum a little bit lower from before um the prior was 1258 and last time, we were at 1260.

So, we are moving a little bit to the left depending on what Sigma p is. if we are very sure then the band will be lower, if we are a little bit less sure it will move a little bit more which is what is happening here. Again, notice q m the mean simply comes from summing up this column Sigma P2 divided by Sigma P1. So, you can see this which is why you get a 10 power 3 factor here so, that comes to 1243.

Similarly, Sigma q Square comes from Sigma P3 divided by Sigma P1. So, that is sitting here and now we have a new maximum a posterior value estimate which is 1240 which is here and a new mean which is 1243 which is different from what we had earlier which been 1258 if I remember that yeah. So, without the prior it was 1258 and now reusing that. Now you can notice what has happened we have zoomed in effectively.

(Refer Slide Time: 54:38)

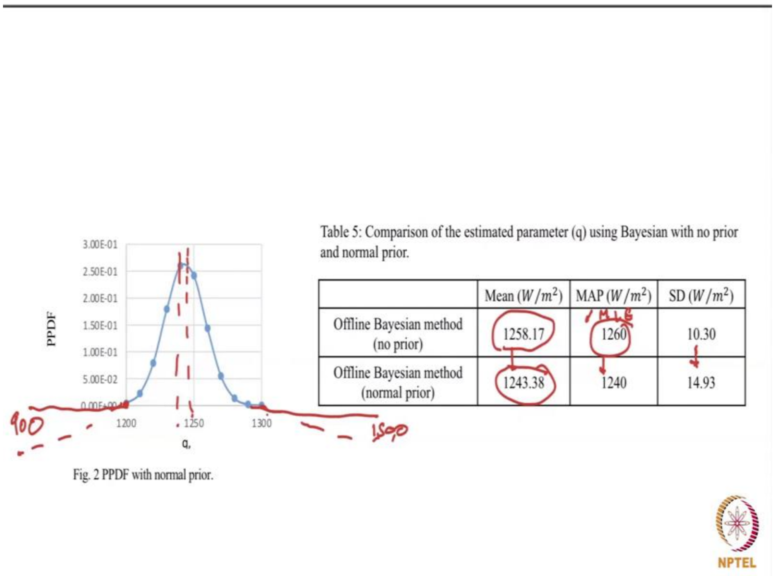


Fig. 2 PPDF with normal prior.

Table 5: Comparison of the estimated parameter (q) using Bayesian with no prior and normal prior.

	Mean (W/m^2)	MAP (W/m^2)	SD (W/m^2)
Offline Bayesian method (no prior)	1258.17	1260	10.30
Offline Bayesian method (normal prior)	1243.38	1240	14.93

Before we had like a strong PDF you know somewhere here is where we were. So, you know I have made this like this really speaking it is flat after this. So, we went till 1500, we went in 900 and now we have zoomed in sort of in computational terms we have done a mesh refinement within this place. So, the maximum occurs somewhere here if you do a more careful simulation you have to be somewhere a little bit before.

You can now compare without any prior the mean was 1258, with the prior basically reusing this prior we got a new estimate which is 1243 with no prior this was basically just the MLE the maximum likelihood. Now using that as prior you come back here the standard deviation was 10.3 the standard deviation actually has become higher. Now why did it become higher? It became higher because the prior we used in this case the prior used was notice this 10 of mu b which is 125.

This is why we became a little bit certain less certain, if we had used a prior of a little bit smaller like 10 or something this would have become more and more narrow. So, this is just an example of an offline Bayesian method. I hope it is clear I will show you the code towards the end of this week. In the next video we will look at the other algorithm which is known as the Metropolis Hastings Markov Chain Monte Carlo which is an online Bayesian approach.

It is sort of just this approach except squeeze to. uh Just one step every step you sort of redo with the information that you have already gone. So, we will see that in the next few videos, thank you.