Meeting Transcript: Variational Bayesian Analysis

[0:00:14] Speaker unknown: I gotta walk upstairs. Now there's
[0:00:21] Speaker 1: on slime three the true cost here is Z, given, not Z, oh, okay, yes, I know it's just going shift the key,
[0:00:36] Speaker unknown: shift in the key under face.
[0:00:40] Speaker 1: Now the oh, that's all fine. The one thing you can probably add at this point, even though we haven't really shown in this book, is that difference between Q optimal and the true loss area should decrease as your data set becomes large.
[0:01:04] Speaker 2: Well, I thought they caught that in there, because if you see the asterisk that ends up taking them being where the KL the is.
[0:01:10] Speaker 1: But my point is, the KL, obviously, is also a function of the size of the data set. So small data sets will pretend have bigger KL then larger does.
[0:01:22] Speaker 2: Curious about something that you say that is that where stochastic modeling helps do it multiple times, because
[0:01:32] Speaker 1: so there's a couple of papers that show variational Bayesian inferences, asymptotically consistent,
[0:01:40] Speaker unknown: which implies that as your sample size increases,
[0:01:44] Speaker unknown: you'll converge to the true distribution,
[0:01:49] Speaker unknown: which also turns into appointments. But skip that next
[0:01:55] Speaker unknown: one black box. It's
[0:02:01] Speaker unknown: not going to be in there. I in there.
[0:02:15] Speaker unknown: Yes, now that's quite
[0:02:19] Speaker 1: good, and if anyone asks you a question, because I suspect
[0:02:24] Speaker unknown: I mean, if Elena asks you a question about science,
[0:02:29] Speaker 1: under certain factorization assumptions, you could add, and because you've said under certain factorization assumptions, these others can be derived in closed form, okay? And you could then say, intuitively, examples where you can find close form team to be examples where you can work out the conditional posterior distribution, okay, because that is that does tend to be the case.
[0:03:01] Speaker unknown: Now,
[0:03:09] Speaker unknown: I Yes,
[0:03:13] Speaker unknown: I don't think you need to show listening.
[0:03:24] Speaker unknown: It helped make the penny drop for me, that's why.
[0:03:33] Speaker unknown: Because, I mean,
[0:03:37] Speaker 2: in one and two and three, I don't think if that's, I don't know if that's, I mean,

I just use that. I think we have been talked about it, but I think that they may have a significance where it's just, for me, it's fine, because

[0:03:47] Speaker 1: you can almost think of them as increasing it's just feel weird. Someone on the audience is going to say, well, what's the point of doing this in one because it's an example. We know the

[0:03:59] Speaker unknown: exact possible

[0:04:02] Speaker unknown: things. Exact cost period,

[0:04:04] Speaker 2: because we're trying to prove that what we're doing is right, we're trying to validate

[0:04:08] Speaker 1: it Yes, but so that's fine. So just be aware that someone in the audience will say, well, in one I know exactly in two different because it's a question of, I don't want to waste my time doing good sampler, because that's the only way I can look it up. You know, I can only do a simulation based example of something I know will converge to the truth if I didn't take enough samples, simulation draws versus variational basics in three you're going to have the issue of having to explain how the hell the variational baby is done, because it's not enclosed. It's not something that everything is

[0:05:02] Speaker 1: know, so just be aware of this. Yes, that this one is not so if someone asks you to reproduce something, in theory, you could work out all the updates in this model. You won't have the background to work out both. Mm. Have the background

[0:05:24] Speaker 2: without what I think I got some errors when you see the actual data. Yeah, I

[0:05:47] Speaker 1: Okay, that's fine. One thing you need to be aware of, that model you have put in here is very specific to model two is very well this is model two, yes, yes. Yes, but the reason I say it is on seven, you don't say that. You need to make that explicit, and you tell them it's model two, yes, okay, yeah, so because, so, I mean, that's fine, maybe a variational principles is conditional independence to the key is effect parameters. We are now going to run through an example corresponding to model two, okay, and then you can go through that, right? And then, if you wanted to, you can add underneath, by the way, if you wanted to do a variational base version of model one, remove the step on tau. You okay, because model one is a special case of model two. Yeah.

[0:06:53] Speaker 2: I saw that in your code. You put zero and everything's good, yeah, for the simple

[0:06:59] Speaker 1: one, now I Yeah. Now, the thing with this is, this is quite hard.

[0:07:16] Speaker 3: The way this is done to pick out the issue, okay.

[0:07:20] Speaker 2: Now the the actual value is Oh, because the issue being the because

[0:07:24] Speaker 1: of the Y axis, yeah, okay, the y axis are not on the same scale.

[0:07:30] Speaker 2: So if you've got oh, those two need to have the same scale, yes, okay, because

[0:07:36] Speaker unknown: you're estimating the same
[0:07:40] Speaker unknown: thing. But you'll notice here,
[0:07:44] Speaker 4: the one on the left goes up from zero to about 1.25
[0:07:52] Speaker unknown: the VB one goes up to 2.5
[0:07:56] Speaker 1: pulse of the Gibbs sampling parameter y axis up to 2.50
[0:08:03] Speaker unknown: okay, yes, yes, yes, then
[0:08:05] Speaker 1: it will be a bit easier to see what has happened. Okay. Because notice at the moment, if someone can't read the axis, they think
[0:08:18] Speaker unknown: green is rhymed.
[0:08:22] Speaker 1: Pink is not right, that Violet is not right, and it gets more not and the orange is not right and the gray is not right, and they'll think it's the small sample size that looks okay per their random effect level. The high one not red. It's actually the other way around. I see what you're saying. So you need to have the y
[0:08:44] Speaker unknown: axis is on top, and I'll make those bolder.
[0:08:48] Speaker 1: Yes, that's why, when I did that code, although I know it's hard to see as well, I tended to put all of them on the one, because of what I was trying to get at is same color, two methods in cases where a lot of random effect lot of observations of the same random effect level, I'm not going to be able to tell the difference. When I don't have many I will be able to tell the difference. Now, the reason we wanted to do random effect model intercept models is quite easy to understand, because we can say sample size, but it's more accurate to say shrinkage. Okay, so because what you're really because the shrinkage is tied to sample size, various components that are tied to rooms that have been very heavily shrunk are going to be examples where various players will do quite badly. Examples where we do not see much shrinkage in the parameters which the variance component is dependent on, we'll do a lot better. So so on eight you need to change the wire So
[0:10:07] Speaker unknown: us systematic. That's fine. That's the data that, yeah,
[0:10:23] Speaker 1: you will obviously don't do very well on model one, yes.
[0:10:28] Speaker unknown: Model Two, yes.
[0:10:36] Speaker 1: And you can see here that that's more or less what you and less what you expected. And the other thing you will know first
[0:10:50] Speaker unknown: is I'm a bit surprised by
[0:10:56] Speaker unknown: those you'll notice the regression
[0:10:59] Speaker unknown: partitions tend to go a lot deeper. To do a lot better,
[0:11:04] Speaker 1: and you'll notice tally does a lot bigger than T. That's not a surprise, because tally has a lot more observation.
[0:11:11] Speaker unknown: Okay, basically is what it has

[0:11:12] Speaker unknown: more observations. Okay,
[0:11:16] Speaker unknown: so the fact it's starting
[0:11:19] Speaker unknown: to creep in here is sort of interesting. Do
[0:11:25] Speaker unknown: is interesting, is worthy.
[0:11:39] Speaker unknown: I changed the change the the equation.
[0:11:43] Speaker 2: Basically, it was the same sort of thing as what, what you did, but it was, end up taking the using the book of the light.
[0:11:52] Speaker 1: Didn't send you code for doing register, no, and it's quite different. Oh, you have sent me one. No, I have not. Okay, I don't think I did.
[0:12:01] Speaker 2: No, I don't think you did. If anything I went, I was winging it, and I think I meant get it wrong,
[0:12:07] Speaker 1: because what I was going to say is, if you've taken the code model two and
[0:12:15] Speaker unknown: done some slight modifications,
[0:12:18] Speaker unknown: that's not going to work.
[0:12:20] Speaker unknown: Okay, I
[0:12:22] Speaker unknown: want them that definitely does not work.
[0:12:26] Speaker 1: The logistic model doesn't have a mean field. Doesn't have a mean field factorization, where everything is known in closed form. And in addition, I you have to calculate if you do knock up something to get around the fact that it's the betas that are clearly not fun. It's these, estimating these, these you can't do in closed form. Okay? And one because the likelihood is no longer normal.
[0:13:09] Speaker unknown: I mean, the there's a couple of ways around it,
[0:13:14] Speaker unknown: so you could do
[0:13:18] Speaker 1: poly gamma data documentation, so rewrite the Bernoulli likelihood as a normal likelihood times a polygamma, then you have to estimate the polygamma terms in your in your algorithm. This is actually implicitly what's done in a very early, novel variation based on just a refreshing like
[0:13:45] Speaker unknown: a gecko, they tend to do quite badly.
[0:13:49] Speaker 1: There's you could do and do things like the black box sort of thing to do better,
[0:13:56] Speaker unknown: because that would do it as well.
[0:13:57] Speaker 2: But isn't that sort of cheating? The whole idea is to understand
[0:14:01] Speaker unknown: this, yes, but my point is modifying the
[0:14:04] Speaker unknown: model to you're not getting
[0:14:06] Speaker 2: what you think. Okay, well, I think that that bears out in the next, next
[0:14:09] Speaker 1: moment, yeah, because some of these patterns here are also not what

[0:14:21] Speaker 1: I mean, that makes this pattern getting a good sample, that makes
[0:14:27] Speaker unknown: a lot wait on m2 Yes, yeah, okay, yeah.
[0:14:29] Speaker 1: But in terms of the timing, because obviously, as you increase the number of random effects, you increase the number of parameters you actually estimate. So the whole thing, but it's
[0:14:41] Speaker unknown: you can actually
[0:14:46] Speaker 1: have in that these, if you look at the last two, there's roughly four times as many parameters, twice as many parameters estimated the last one,
[0:15:01] Speaker unknown: between 50 and 100 Yeah, yes, okay.
[0:15:03] Speaker 1: I mean, because I know there's a couple of other parameters, but roughly double. So you can see it's roughly the GIF samples got a quadratic cost. Yes, yes, yeah. We, as you can see, roughly doubling. You've got a bit more than four times 20 to 50. You can sort of see at the beginning, at the beginning, you tend not to notice it too much, but yeah, as I said, and the other thing is the give samples for the
[0:15:43] Speaker 2: Okay, so that's it. I mean it just, I never understood why it was doing so so small,
[0:15:54] Speaker unknown: because you've got to be very, very is that
[0:16:01] Speaker unknown: expected for into
[0:16:07] Speaker 1: Yeah, and I was cut out singing. I was cut out in free cardboard talk. I don't want
[0:16:17] Speaker 2: to do that. I'll tell you what I want to do. I want to say that I screwed up. I want to be honest about that's fine.
[0:16:27] Speaker 1: You can say you started. You can say you started on I am afraid, but we didn't get it time to finish, and I need to fix the code so we've started an
[0:16:36] Speaker unknown: initial team. Right? This one is the weirdest.
[0:16:48] Speaker unknown: Well, I mean, in tahini two, that's
[0:16:53] Speaker 1: what I would expect. But again, the higher up was passive. This is doing this because
[0:17:00] Speaker 2: of all those wrong but what I didn't understand here is that the two formula that the the m2 and m3
[0:17:08] Speaker unknown: acted the same, but on the on the other method,
[0:17:13] Speaker 2: one method on top, another method on it, but They They basically follow the same pattern. And now again, I did this wrong, but still, well,
[0:17:27] Speaker unknown: that doesn't actually surprise me as much as
[0:17:32] Speaker 1: you say, because
[0:17:39] Speaker unknown: you've got to remember

[0:17:43] Speaker 1: the variational phase. You know the variational base algorithm is not it's better. The solution is still iterative. So it's not just that more groups make longer you've also got the question of, how many iterations does it take to converge depending on the size of the group? So with this example here, what you might be picking up is it needs a lot of iterations. Whereas once you've got more groups, it tends to get into the right position in fewer iterations, but doing it in single iteration takes longer. So this sort of bowlish pattern you're seeing here might not be as much of a surprise if you work out how many iterations, that's fine, as I

[0:18:38] Speaker 4: said, the slide

[0:18:45] Speaker unknown: that, yes, I gotta fix

[0:18:47] Speaker unknown: it. And

[0:18:50] Speaker 4: as I said, you do need to be we have a hierarchical that goes

[0:18:56] Speaker unknown: the hierarchical model, the logistic

[0:19:00] Speaker 1: one started it, but it needs a lot more work. And make that clear that that's model two, you're referring Yes.

[0:19:19] Speaker unknown: Would model three be the same but different equations.

[0:19:25] Speaker 3: It's worse than that. Okay, so

[0:19:32] Speaker unknown: the issue with model three is the

[0:19:36] Speaker unknown: likelihood function,

[0:19:41] Speaker 1: so the livelihood function for y does not lend itself to a closed form of beta. Well, not, not unless you use polygamma Double augmentation. So you would get quite similar equations to this. If you use polygamous augmentation, but then add in another set of parameters, you have to estimate which are these Polya gamma realizations. You have to do it for every single observation.

[0:20:13] Speaker unknown: You have to do one. You have to do it for every

[0:20:15] Speaker unknown: single you have to estimate it every

[0:20:18] Speaker 1: year, okay, yeah, because it's, it's data point specific. And so what it means is so EA you would get so that we know tau e coming through here, but you have an omega coming you have the expectation of a vector of omega coming through here. You also have to work out the mirrorings, I believe, because these omegas also will appear in here. So you would tell E's get replaced with lots and lots and lots of omegas. And you then have to estimate these omega and I can't remember, you just estimate omega, or you have to estimate omega squared and then square root.

[0:21:14] Speaker 2: What? Why would? I don't understand what the difference would be. I mean, you're still getting the same because

[0:21:20] Speaker unknown: it's because you take the expectation first.

[0:21:24] Speaker 1: So the expectation of each squared not the expectation. It's not the same as

the expectation of these being squeezed. Oh, okay, okay, so square rooting, the expectation of each set was mean squared is the expectation of e the expectation of each squared being square rooted is not the expectation. Um, but you would need, if you want to understand that, you would also have to read, you'd have to read the paper by Paulson and gaugmentation explaining how you can

[0:21:57] Speaker unknown: rewrite binomial life.

[0:21:59] Speaker 2: I would like that note that paper, you don't have to tell me. Now, if you just give me a general aspect in the email that I can just do a search, I'll do well, I've got

[0:22:09] Speaker unknown: a couple. Okay, thank you.

[0:22:13] Speaker 1: So in that, what? And in addition, I know that teams, not from what I gather in literature, and the logistic regression case, it doesn't perform that well because essentially putting Napoleon gathers, essentially by partitioning the livelihood, you're doing another round of factorizations that introduce more shrinkage being moved. I mean, the other way around it is. The other way around it is to assume a fixed form, assume the posterior for beta is normal, and then do something that looks

[0:22:59] Speaker 1: quite loud, the estimation of GL ends. But you have to work out certain expectations which aren't known.

[0:23:12] Speaker 4: That's fine, but that, as I said, that just means rescaling to make point, and you need to

[0:23:22] Speaker unknown: make the bold the axis, yes, and we can

[0:23:27] Speaker 1: talk about that while in this example, if we sample size a more general way, this is fine, but please, if you're going To leave model three and emphasize it's not correct, yes, and the same. That's fine, as I said, this one check

[0:23:50] Speaker unknown: how many iterations it takes, what you're seeing as a

[0:23:54] Speaker 1: combination increasing, however, the size makes doing a single iteration take longer, but meaning some very small models it requires. Okay,

[0:24:13] Speaker unknown: so that's that I have looked through these.

[0:24:24] Speaker 2: Well, the thing about it is, this is where I this is where I got the data, and it ended up taking in where I you can see where I ended up thinking of changing your actor. Yeah, it's so one

[0:24:34] Speaker unknown: I'm going to bring up is the logistic one.

[0:24:55] Speaker unknown: They wasn't done.

[0:25:08] Speaker unknown: My knowledge at this point is is which needs to be?

[0:25:37] Speaker unknown: Are these good sample of that line that one doesn't understand? Because

[0:25:59] Speaker 4: this means are always changing, whereas the for a jet, sometimes you don't train, I'll need to look at the good times and in my

[0:26:34] Speaker 3: out and hear that either you
[0:26:46] Speaker unknown: that looks too. That looks
[0:26:50] Speaker unknown: GL in like, not VP, not
[0:26:54] Speaker unknown: based. So these nines here,
[0:26:58] Speaker unknown: that's not what you should
[0:27:01] Speaker 1: get, because what you're doing there is not the expectation of P, but you're doing the function of an expectation of
[0:27:25] Speaker 1: and also, and this is coming from the proposal to not arrive. That's why these, this is why these are often weird, because they they should. They might not sit directly on top, but they should be smooth. They should be smooth. Bell shaped curves of the bead is in their mouth, which is telling you,
[0:27:55] Speaker unknown: and the same Yes. So yeah, it's
[0:28:00] Speaker 4: the code, um. You realize you would need to
[0:28:10] Speaker unknown: manage to run that properly.
[0:28:11] Speaker 2: Well, you are going to give me a paper, at least, because we started in that
[0:28:16] Speaker 1: area, yes, but being warm, that's fairly mathematical, pretty simple. Any other questions,
[0:28:25] Speaker 2: would you allow me to go to the likelihood class, undergraduate class?
[0:28:29] Speaker unknown: Well, I
[0:28:32] Speaker unknown: either call escorting. I tried doing that.
[0:28:42] Speaker unknown: I was dying to say, do you need special permission to enroll in
[0:28:48] Speaker 2: the class? Well, I don't know. I'm just asking for it in advance.
[0:28:53] Speaker unknown: I really don't, I mean, it's nice,
[0:28:57] Speaker unknown: I really don't. If there is an issue with your enrollment, actually,
[0:29:12] Speaker unknown: what maths have you done at
[0:29:16] Speaker unknown: university here or anywhere?
[0:29:17] Speaker 2: Oh, calculus, differential equations. Okay, these maximums.
[0:29:22] Speaker unknown: Maximum. The reason I ask that is
[0:29:26] Speaker unknown: step 213, has a 100
[0:29:31] Speaker 1: level mass required. I did do this 40 years ago. Yes, so you need to be warmed. There is an expectation about an expectation about you know how to integrate, you know how to debrief? Yes, in fact, it's usually one of the things that shows up with the student cohort in 213, that the students who've only got 100 level maths,
[0:30:01] Speaker unknown: they can they will pass, but they will struggle.
[0:30:05] Speaker 1: The students who are taking the course but have done at least some 200 or 300 level maths, they do a lot better. So be warned,

[0:30:18] Speaker unknown: it's the maths background.
[0:30:21] Speaker 2: Should I take a particular math class in advance of this? Well, I don't mind taking extra classes. I don't mind delaying my degree. So
[0:30:33] Speaker unknown: prerequisite, post fact, 213, is math 102,
[0:30:41] Speaker 1: now you have clearly done your problem with that, even though it was quite some time ago. Okay, so I don't know if you would be in it might be a case of talking to the teaching team,
[0:30:57] Speaker unknown: but what you were saying was this the 200 level that
[0:30:59] Speaker 1: I'm missing. Yeah. No, no, this is a 200 so step 200 is a 200 level status. Course, to get into it, you have to have completed 100 level math. Okay, so and I, the thing I was adding is what you tend to find it. But happens in practices, if you complete 100 level math. You will be in the students who have just done 100 level math and no more will struggle. They will pass. But they will struggle because the students who have gone and done 200 or 300 math papers, there's, you know, they're doing it later in the degree, they do a
[0:31:40] Speaker 2: lot either, can I see the curriculum of the 200 to see if I end up taking, not the curriculum, but can I see some of the well, the because then I can end up taking annoying Is this beyond where I where I
[0:31:52] Speaker unknown: might understand you?
[0:32:14] Speaker 1: So obviously it varies with time. So in the beginning, it's reminding people about random variables. In fact, some of this stuff might seem a bit familiar from the start to be wonderful because Elaine talks often introduces conditional
[0:32:33] Speaker unknown: probability, so you get distribution functions.
[0:32:41] Speaker unknown: Probability, little example reminding people what a gainsay function
[0:32:58] Speaker 1: is being I go on about specific distributions, which, again, should be somewhat familiar from
[0:33:06] Speaker unknown: step two on, the three end moments and calculating
[0:33:16] Speaker unknown: quantities. It gets worse on the
[0:33:20] Speaker unknown: second half. So more distribution. So there's quite a collection of
[0:33:28] Speaker unknown: distribution and then generating functions.
[0:33:36] Speaker unknown: Gamma distribution process
[0:33:41] Speaker unknown: is chi squared. And
[0:33:55] Speaker 2: variances. So far, I haven't seen anything that I can't end up taking any warming up on.
[0:33:59] Speaker 1: No So, as I said, you might not need to do the 100 level math, but it's
[0:34:05] Speaker unknown: just if you're coming out of, you know, if You're full
[0:34:13] Speaker unknown: correlation dependence, then we'll get

[0:34:21] Speaker unknown: into central limit. Theorem, distribution of the mean limiting probability, convergence,
[0:34:27] Speaker unknown: inequalities, convergence
[0:34:30] Speaker unknown: of quite a bit mean or large numbers.
[0:34:36] Speaker unknown: That's what I don't know, properties of a chi squared Quite a theorem, which I
[0:34:45] Speaker unknown: tend to do fast then
[0:35:00] Speaker 1: the properties of estimated sampling distributions, convolutions of random variable,
[0:35:14] Speaker unknown: uniform I remind you that I was into limit theorem.
[0:35:19] Speaker unknown: Some examples bias into the era. Comparable
[0:35:26] Speaker 1: moment you'll remember maximum likelihood. This is because Elaine did it then during the first half of 314,
[0:35:35] Speaker unknown: IEP, so that gets done to
[0:35:58] Speaker 1: equivalent to removing there is a couple of lectures on these and statistics, but it's very much the much doing examples. This
[0:36:17] Speaker 1: is one that students don't deny. So as I see, it might be a case of fish out very old notes you have on that, and you should be fine. But being warm, there is an expectation you can integrate a
[0:36:31] Speaker 2: different machine. Okay, I've got, I gotta warm up on that as well. I mean, you know, it's like calculus. When I first class, I did with calculus back for five years ago. You know it was, it was familiar, and then it was like, oh, yeah, okay. And, but it wasn't something in which I was right off the top of my head, which is what you where people are that just ended up taking having a class. Yeah, exactly. And it's like I said before, when we are end up doing our presentations, and you can see what one gentleman was doing. I still haven't figured out that how you ended up doing that. But the thing is, it's that familiarity that I would like to be able to at least try.
[0:37:10] Speaker unknown: Okay, Mark is covered. Okay, yes,
[0:37:18] Speaker 1: my old happy. I'm happy to look over another version of your drawing before Friday.
[0:37:22] Speaker unknown: Draft before Friday, which is not
[0:37:27] Speaker 2: Friday, it's Thursday. Tell me it's Friday. I would love that
[0:37:31] Speaker 1: it's Friday. Friday presentation on Friday, the report to June, next Tuesday.
[0:37:39] Speaker unknown: That means I have, I have two days to end up taking and
[0:37:42] Speaker 1: getting this right at unit sees if you want your slides to be pre loaded onto the computer in the room or just to have a backup, please email the slides before 5pm on Thursday.

Okay, don't have to do that. If you wish to edit or finalize afterwards. It is fine, but you will be responsible to bring them up over the slides to the room on Friday. Okay, cut down at the file size is large, greater than 20 megabytes. You can use a cloud service like Google Drive,

[0:38:11] Speaker unknown: okay, it's, I think it's 8 million

[0:38:13] Speaker 1: Well, presenting, presenting, you can either use a room computer or bring your own laptop via HDMI or BGA.

[0:38:21] Speaker unknown: They will supply the cable?

[0:38:23] Speaker unknown: Yes, because it'll be

[0:38:26] Speaker 2: well, what I want to do is, at least tomorrow