EXTENSION AND CORRECTION TO THE INCONSISTENCY EXAMPLE

As I described it in the lecture, the inconsistency example was incorrect. I said θ indexed a Gaussian location parameter, e.g. with the model having X_t i.i.d. $N(\theta, 1)$. This doesn't work. Ghosh and Ramamoorthi's version of the example, on p.31-32 of their book, makes X_t i.i.d. $U(0, \theta)$, and this particular family of distributions is important to the example.

With the prior restricting the distribution of X_t to $U(0, \theta)$ with $\theta \in (0, 1) \cup (2, 3)$, $\theta = 1$ is not in the K-L support of the prior, and this is the condition from the Schwartz theorem that fails in the example. Recall that the KL support of the prior is the collection of θ values such that the prior probability $\Pi(U)$ of the set U of all distributions within KL distance ε of θ is positive, for any $\varepsilon > 0$. Recall here also that we base the KL distance in this definition at θ itself, so with $\theta = 1$ and another density f, the set U is of the form

$$\int_0^1 -\log f(x)\, dx < \varepsilon\,.$$

Any f that is zero over an interval where the $U(0, 1)$ density is positive (i.e. over any interval in (0,1)) is at infinite KL distance from the $\theta = 1$ density. That means the only uniform densities at finite distance from the $\theta = 1$ density that have positive prior density are those in (2,3). But those are all at KL distance at least $\log(2)$, since $\theta = 2$ is the closest of them. So sets of $U(0, \theta)$ distributions at distance less than ε from the $\theta = 1$ distribution all have $\theta > 1$, and if ε is small they do not include $\theta \geq 2$. But the prior in the example has positive probability only on uniforms, and zero probability on $\theta \in (1, 2)$. So $\theta = 1$ is not in the KL support.

Note that this argument did not use the particular form of the prior density, only the fact that it was concentrated on distributions $U(0, \theta)$ with $\theta \in (0, 1) \cup (2, 3)$. But with a prior density on θ uniform over $(0, 1) \cup (2, 3)$ instead of with the special $\exp(-1/(1-\theta)^2)$ of the example, the posterior would be consistent. This makes the point that the Schwartz conditions are sufficient for consistency, not necessary.

On the other hand, if $X_t \sim N(\theta, 1)$ is the model, $\theta = 1$ is in the KL support, which tells us, without having to evaluate nasty integrals, that the posterior is consistent in this case.