Sup doc here are comments on (4)...

**SECTION 2.1**

**Section 2.1.1...**  
All is well in first two paragraphs.

**Last paragraph: your intuition is correct but how u wrote things should be slightly changed**.

**U r confusing operations on events, which are unions and intersections, with operations on probabilities**.

**Unions between disjoint events map to sums of their probabilities**.

*X=a Union Y=b, if these are disjoint events, then the probability of X=a Union Y=b is equivalent to P(X=a) + P(Y=b)*

 But u r taking "sums" of unions.  You are conceptually correct but u see what I mean?  
  
u r trying to say things like we need to say that "the event that X=k and Y=v"...  U write that mathematically with the curly brace  
  
\{ \{X=k\_1\} \cap \{Y=v\_1\}  \} \cup ...  
  
Cap means intersection and cup means union.  Curley brace wrapped around random variables doing shit means u r explicitly talking about an associated event.  
  
**It might be better to define A\_j and B\_k once and then do set operations on them**  
So u see how to change that equation right above section 2.2?

**The last thing also should be a union rather than a sum, and that union should not go all the way to Omega (the analogue of "1"),**

**rather all of that should go to just the event that X equals k\_1**.

*This makes sense because the event that X=k\_1 does not necessarily represent the entire sample space.*

**Conditional probabilities add to 1 but note u only have intersections there, u don't have the denominator in the law of conditional probability**.

*In other words law of conditional probability: P(A|B)=[P(A intersect B)]/P(B)*

Thus that union will amount to the event that X equals k\_1.  or equivalently the associated probability of that union will be P( X=k\_1 )

*Makes sense. The event that X=k\_1 is the whole event and we are just chopping it up into intersections between X and various values of Y, but X=k\_1 incorporates the entire event*  
  
So I want u to carefully think through that last paragraph in section 2.1 carefully.

**SECTION 2.2**  
  
**In section 2.2 one intuitive part u r missing is the issue of "Delta**".

**I want u to incorporate the definition of the density of a continuous RV as a derivative but rewrite it in terms of an interval of length Delta**.

What is the probability of a RV lying in an interval containing the point 50, where the length of the interval is Delta?  
  
I also didn't understand what u meant when u say we "probe it to a finite number of values".  I think u r trying to say that we only have finite precision in the measurement.  I see how u r thinking but that perhaps is not the best way to think of it.  The issue is that your observation can be exact so u can know EXACTLY what value Y took.  But asking a question about that associate probability, in some sense, is vacuous for a cts RV.  The probability that Y takes on exactly any one point is 0.  I know it is counter intuitive because Y DOES take on exactly one point.  
  
**A better statistical question to ask is, what is the probability that Y lies in an interval of length Delta centered at 0.2?  And compare that to the probability that Y lies in an interval of that same length Delta centered at 0.79?**  
  
That is the "right" way to think statistically about cts RVs and in particular how to compare the odds of Y lying in the vicinity of 0.79 vs the odds of Y lying in the vicinity of 0.2.  If u asked the question about EXACTLY 0.2 and 0.79, u have 0 vs 0.  But ask about them lying in two small intervals, both of length Delta, centers at 0.2 and 0.79, and u get relevant answers.  
  
There is nothing wrong u said in the first part of section 2.2 preceding section 2.2.1 but I didn't understand what u were trying to say?  **And the punch line is that a sum over probabilities involving Y taking a value v become integrals.**

**2.2.1**  
Everything in section 2.2.1 looks fine.  I would change that last math stuff in that last paragraph to involve the random variable Y instead of X.  Just to be consistent w HMM notation because Y is the one who is cts.  Remember as of now, X will still be discrete.

*Makes sense. That is done*  
  
The good news is that the conditional PMF of X given the Y observations for the filter is still just a PMF!  So the calculation will be very very very easy once u get it conceptually... Also look back to one of the last few homeworks where I asked questions about conditional PMF of a discrete RV X given a cts RV Y equals v.  
  
Go back and make sure u understand that..  Then this becomes super super easy...