ISTA 410/510 Homework V

For contribution to the final grade, due dates, current late policy, and instructions for handing the assignment in, see the assignment web page.

Please create a PDF document with your answers and/or the results of any programs that you write. You should also hand in your programs.

Questions marked by * are required for grad students only. They count as extra problems for undergraduates. (Unlike previous assignments, the (*) are not much harder than the non (*) assignments.

Undergraduates need to do 10 points worth for full marks, and graduate students need to 16 points (all questions). While there are no alternative questions for grad students in the current version of this assignment, you could propose your own. Email Kobus if you have an idea.

Any non-challenge problem can be replaced by a challenge problem; please make it clear that this is what you are doing (e.g., for a required problem you could answer "see optional problem #3"). The point here is to enable students to avoid problems that they feel are not instructive.

Extra problems (please indicated in your answer when you are doing an extra problem) are eligible for modest extra credit. The maximum score for an assignment will be capped at 120%. The maximum score for all assignments taken together is capped at 65/60.

For simplicity, problems are generally all worth the same, except ones marked by "+" that are expected to substantively more time consuming, and are worth double (or two no "+" problems). Additional "+"s scale linearly.

Sex in flatland

The sex of the magnetic-quad-limb (MQL) creatures in flatland is hard to tell. Let's write a computer program to help them out.

There are twice as many males as females. The bodies of males are circular and that of the females are hexagonal. These shapes are easily confused as specified below. The color of the bodies are gray. The bodies of the MQL creatures always align with the external (known) magnetic field. This means that there is a well defined coordinate system for the angles of each limb leaving their bodies. MQLs have two arms and two legs, each having an inner segment attached to the body, and an outer segment attached to the inner segment. The inner arm segments (IAS) are golden, the outer arm segments (OAS) are purple, the inner leg segments (ILS) are red, and the outer leg segments (OLS) are green.

The colors can be reliably detected, and it is always clear which of the two green blobs (for example) are the left OLS (LOLS) and which is the right OLS (ROLS). Notice however, that the colors do not tell you about the sex of the MQL.

Other than the unreliable body detection, evidence for the sex of the MQLs comes from the limb angles. However, because the MQLs are quite furry, angles are also unreliably measured.

MQLs are symmetric, so the right inner arm angle (RIAA) and left inner arm angle (LIAA) sum to 180 degrees. Similarly, the right inner leg angle (RILA) and left inner leg angle (LILA) sum to 540 degrees. Further, RIAA is between 0 and 90 degrees, and RILA is between 270 and 360 degrees. Note that these determine the LIAA and LILA respectively. The distributions of RIAA and RILA are different for males and females. Finally, the outer limb angles, measured with respect to the inner limb angles (see figure) are independent of the inner limb angles. The distributions are the same for arm outer segments and leg outer segments, and they are the same for left and right. This distribution does not depend on sex (until question 9).

We will consider that angles are discretized into 6 equal sections for each quadrant under consideration in 15 degree increments.

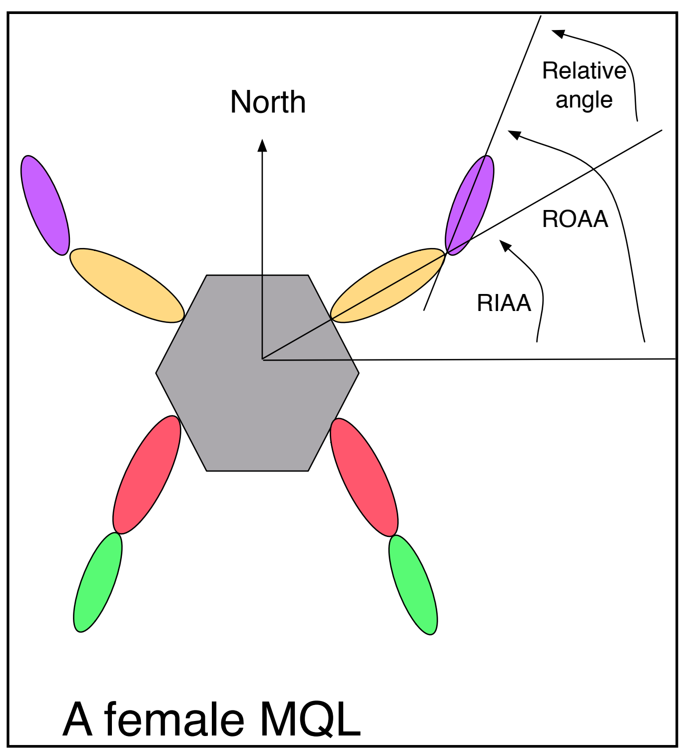
Females have RIAA distributed over the 6 sections as 10,20,30,40,50,60.

Males have RIAA distributed in proportion to: 10, 20, 40, 40, 20,10.

Females have RILA distributed over the 6 sections as 10,20,40,40,20,10.

Males have RILA distributed in proportion to: 60, 50, 40, 30, 20,10.

Both males and females have both outer segments distributed from -45 to 45 degrees around the angle of the inner limb that they are attached to. Again we discretize into 6 angular regions in 15 degree increments. The relative angles are distributed in proportion to: 30,20,10,10,20,30. (Yes, the are bent).



If it is helpful, you can assume that all angles are multiples of 15 degrees.

Note that the model for the outer segments are defined in terms of the relative angle, but what is measured is the absolute angle. It is important not to confuse the difference of the measured inner and outer angles with the true relative angle—the former does not have the statistics specified above. In particular, note that you can observe an angle that appears to be impossible.

For each of the 6 angular "bins," for the true value, the angle that is actually measured has 20% chance of being measured in the previously numbered bin (if there is one), 10% chance of being measures as coming from the bin numbered two less (if there is one), and similarly, it has a 20% chance as appearing to come from the bin in the right (if there is on) and 10% chance of appearing to come from the bin numbered two more (if there is one). This measurement error is independent for each measurement made. Hence, although we know that MQLs are symmetric, they may not appear to be.

Finally, a circular body has a 35% chance of being detected as a hexagon, and a hexagonal body has a 50% chance of being detected as circle.

We will assume that we never see anything that is impossible, which is a safe assumption because we will work with synthetic data.

Enough foreplay. Now for some questions!

1. Produce the table p(measured_angle | angle). Do you think that the variance of the measurements will be the same for each angle?

2. Provide a graphical model for the hidden variables and the observable variables.

3. Provide a probabilistic formula for the hidden variables and the observable variables.

4. Write a program to generate samples from the probability distribution.

5. (+) Develop a simple way to draw pictures indicating generated male and female MQLs[1]. Provide 10 samples of each, and illustrate them in two representations. 1) The idealized individual; and 2) The observed appearance is (i.e., (1) with sampled measurement noise). Can you tell them a part reliably from (1)? How about (2)?

6. (+++) Implement a sum-product algorithm to determine the sex of the MQL based on the observed data. Your algorithm should provide the posterior probabilities for male versus female. Now suppose that you have to declare whether an MQL is male versus female. The natural thing to do is to provide the maximally probability one, as we do not have a specific risk function in mind. Test the accuracy of doing so on sufficient generated data so that the accuracy estimate is stable.

7. (*) Generate a plot indicating the percentage of males as a function of probability of males provided by your program, in suitable increments (e.g., 0-10% chance of males, 10-20%, etc).

8. (*) (+++) Implement a max-sum algorithm to determine the maximal configuration of the MQL hidden variables. Compute this for at least 10 generated measurement sets, and plot the configuration together with the truth, and the measured data. Use this to determine MQL sex by simply consulting the sex variable in the maximal configuration, and study the performance in the same way that you did for sum-product approach. Comment on the difference or lack thereof.

---

[1] Code will be provided to help generating images for those who do not want to do their own graphics.

9. (*) Suppose that the outer limb angle distributions are different for males versus females. Does this change your graphical model in (2)? If so, provide an updated one. Ditto for the probability distribution you provided for (3).