Questions

1. Suppose we have a likelihood function \( p(I|S) \) and a prior \( p(S) \), and that both have a Gaussian shape. The prior, in particular, will be a Gaussian probability function. Further suppose we have an instance \( I = i \) and we can compute the likelihood function. How can we combine the likelihood function and prior to estimate the value of \( S \) that maximizes the posterior?

   Hint: the derivation is essentially the same as given at the beginning of the lecture for combining two likelihood functions in the case of conditional independence.

2. What if we have two likelihood functions and a prior, for example, we have texture and stereo cues that give us (conditionally) independent likelihoods of surface slants, and we also have a prior? How do we combine the weights so that we can compute a maximum posterior probability of \( S \)? Assume conditional independence of the two likelihood functions, so:

\[
p(S|I_1, I_2) = \frac{p(I_1|S)p(I_2|S)p(S)}{p(I_1, I_2)}.
\]

3. One might expect the prior \( p(x,y,d) \) on binocular disparities to have its peak at \( d = 0 \) for image positions near the center of the fovea(s), i.e. near \( (x,y) = (0,0) \), since the eyes tend to verge so that the center of fovea has zero disparity.

   Would you expect the prior to also have its peak at \( d = 0 \) for other \( (x,y) \) values? In particular, would you expect the prior for points above the line of sight to have its peak at \( d = 0 \). Similarly what about below the line of sight?

   Hint: Recall that there is a prior for floor orientations over ceilings.

4. (a) The two images on the left below look weird, but for different reasons. Why?

   (b) The drawing on the right (due to Roger Shephard (?)) shows two trapezoids. The 2D shape and size of these two trapezoids appear to be different: the one on the left looks longer and skinnier than the one on the right. But in fact they are the same. How would you explain this illusion in terms of likelihoods or priors.
5. [ADDED: March 12.]

(a) Describe a psychophysical experiment to measure how accurately people perceive the orientation of a line in an image. Mention what happens on each trial, what the task is, how the experimenter determines the just noticeable differences (JND), and what is plotted to illustrate the result.

(b) What would a maximum likelihood model predict about how the plot changes as the image noise is increased? Briefly explain why.

(c) Natural visual environments tend to have more vertical and horizontal structures than diagonal structure. For example, trees and buildings are vertical and objects lying on the ground are horizontal. These prior probabilities suggest a Bayesian model of image line orientation perception. What specific predictions would such a Bayesian model make about perceived line orientations?
Solutions

1. Let the likelihood function \( p(I = i|S) \) be a Gaussian with mean \( S = s_1 \) and variance \( \sigma_1^2 \). Let the prior be a Gaussian with mean \( s_p \) and and variance \( \sigma_p^2 \). The solution is exactly the same as on page 2 of lecture 13 where we maximized \( p(I_1|S)p(I_2|S) \). In the current problem, we want to maximize

\[
p(I|S)p(S) = \frac{a_1}{2\pi\sigma_p} e^{-\frac{(s-s_1)^2}{2\sigma_1^2}} e^{-\frac{(s-s_p)^2}{2\sigma_p^2}}.
\]

You can crank through the same derivation, or just plug in the new names of the variables. You get:

\[
w_1 = \frac{\sigma_p^2}{\sigma_1^2 + \sigma_p^2} \quad w_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_p^2}.
\]

2. We now want to maximize

\[
p(I_1|S) p(I_2|S) p(S) = e^{-\frac{(s-s_1)^2}{2\sigma_1^2}} e^{-\frac{(s-s_2)^2}{2\sigma_2^2}} e^{-\frac{(s-s_p)^2}{2\sigma_p^2}}
\]

so we want to minimize

\[
\frac{(s-s_1)^2}{2\sigma_1^2} + \frac{(s-s_2)^2}{2\sigma_2^2} + \frac{(s-s_p)^2}{2\sigma_p^2}.
\]

Take the derivative with respect to \( s \) and set it to 0 gives

\[
\frac{s_1 - s}{\sigma_1^2} + \frac{s_2 - s}{\sigma_2^2} + \frac{s_p - s}{\sigma_p^2} = 0
\]

and so

\[
s = \left( \frac{s_1}{\sigma_1^2} + \frac{s_2}{\sigma_2^2} + \frac{s_p}{\sigma_p^2} \right) / \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_p^2} \right)
\]

This is the estimate of the maximum of the posterior. Note that this is of the form

\[
s = \sum_{i=1}^{3} w_is_i
\]

where

\[
w_i = \sigma_i^{-2} / \sum_i \sigma_i^{-2}.
\]

3. The hint is that there is a prior for floors over ceilings. This suggests there is a prior for points above the line of sight to be further away than points below the line of sight. For the disparity prior to be consistent with this slant prior, points above the line of sight should have a disparity prior that peaks at a negative value (not at \( d = 0 \)). Similarly, points below the line of sight should have a disparity prior that peaks at positive value.

4. (a) The image on the left is an upright face illuminated from below and the one on the right is an inverted face illuminated from above (i.e. just the flipped image of the one on the left). The image on the left looks weird because we are used to upright faces that
are illuminated from above. The illumination from below creates shading patterns that are consistent with flips in the surface depth, which are inconsistent with the 3D shapes of faces e.g. the local shading may suggest a local valley (if the patch has a dark spot above a bright spot) whereas the configuration of the features of the face suggest the local region is a hill.

For the image on the right, we perceive an inverted face because of the configuration of the eyes and mouth. This itself looks a bit weird because normally faces are upright. That said, the surface still does have an overall convexity and is perceived as such. Moreover, the local shading is consistent with the overall convexity of the surface. Local convex regions are shaded as local convex regions should be when illuminated from above, namely the bright region is above the dark region.

(b) If you remove the table legs from the drawing and just have the two shapes, then they look identical except they are rotated. So the table legs make a huge difference. They provide occlusion cues (and more generally pictorial cues) that suggest the shapes are the tops of table and are standing on the ground. The prior for viewpoint from above helps out with this percept. Because the legs suggest that the table tops are slanted away in 3D (tilt of 90 degrees), the shapes must be foreshortened in the Y direction. This implies the shape on the left is longer than it appears and the shape on the right is wider than it appears.

Note that if you flip the image so that the legs are above the shapes, then the tabletop percept disappears, and the shapes no longer appear to be different. The tabletop percept presumably disappears because we rarely see tables pinned to the ceiling. And the occlusion information that says the legs are perpendicular to the table is weaker, for similar reasons.

5. (a) Each trial could consist of two images, each containing a line, and the subject could be asked to judge which image had the more vertical line. For each reference orientation, a psychometric function could be plotted as a function of the test orientation (or the difference in the test and reference orientations). Only orientations very near the test orientation would need to be considered as the task would be trivial for large orientation differences. The JND could be the orientation difference between reference and test that is needed to get 75 % correct. One would plot the JNDs as a function of the reference orientation.

(b) For each reference orientation, the JNDs would rise. That is, you would need greater differences in the line orientations in order to obtain 75 % correct.

(c) The Bayesian model would bias the perceived line orientations toward vertical and horizontal. Specifically, it predicts that any orientation that is less than 45 degrees from vertical would be perceived as closer to vertical than it really is, and any orientation less than 45 degrees from horizontal would be perceived as closer to horizontal than it really is.

As noise is increased, the bias toward vertical and horizontal would increase since the weight from the noisy image would be lowered relative to the weight from the prior.