Today I will revisit some of the ideas that I introduced last lecture, and I will try to pose these ideas in terms of probabilities.

**Motivation**

To put you in the mood, let’s reconsider some of the vision tasks we discussed last class. First let’s discuss the task of detecting an increment in intensity $\Delta I$ in the center of a uniform intensity field $I$. How can we think about this in terms of probabilities?

Suppose that the observer has some uncertainty in the intensity of both the center intensity $I_0 + \Delta I$ and the background $I_0$. This uncertainty can be due to noise in the monitor or to the noise in the visual system. If the noise level is large relative to $\Delta I$ then it will be difficult to say if there is indeed an increment $\Delta I$ or not.

The term ‘likelihood’ has a formal definition which I will review a bit later. For now, let us take the likelihood intuitively to be proportional to the subjective probability that the background has value $I_0$ and the center square has value $I_0 + \Delta I$, considering only your uncertainty that is due to the noise.

As another example, consider the orientation of a 2D sinusoid image with some additive noise. If the contrast of the sinusoid is extremely low, then you will not be able to make out the pattern of the sinusoid and all orientations in the image will be equally, based on what you can observe in the image. As you increase the contrast of the sinusoid, it will start to become visible in the image and the likelihood will be elevated that the correct orientation. Eventually, when the contrast of the sinusoid is very high, there will be an elevated likelihood at only a small range of orientations near the true one.

A third example I discussed in the lecture is image motion. I introduced a different image motion stimulus which consists of random dots moving in 2D. (This was different from the example considered in the motion constraint equation, where I assumed we could take image derivatives.) Each dot either moves with some fixed image velocity $(v_x, v_y)$ or else it moves with some random velocity drawn from a distribution. For this stimulus, the random distribution is fixed, and we can consider different percentage of dots – called the motion coherence – that move with the given velocity $(v_x, v_y)$.

On the next page I sketch out informally the likelihood function for this case. For 100 % coherence, the likelihood will be concentrated around the true velocity. There will be some spread because of noise in the visual system, for example. For 50 % coherence, there will be more of a spread in the likelihood because half the velocities are in a random direction, but again centered
around the correct velocity. If one were trying to write down a formal model, one might argue that
the center of the distribution should be shifted toward the origin. I am not going to argue against
that – it would depend on the model. My point here is just to sketch out the basic idea, namely
the there is a spread in the likelihood as the noise increases. For 0 % coherence, there would be a
large spread in the likelihood, and I have sketched it so that it is centered at the origin.

The next example I mentioned in class is for depth from binocular disparities. (Figure omitted.)
If the stimulus is a random dot stereogram with a center square protruding from a background then
the likelihood function will be similar to the one on the previous page for \( I \) and \( I + \Delta I \) except that
now it will be disparity of the background and the disparity of the center.

The final example for now is shape from texture. Take a class of stimuli in which the texture
is generated by random shaped and sized ellipsoids on a plane. The noise of the texture is defined
by the distributions of ellipse shapes and sizes. For a single image ellipse, one cannot say whether
it comes from an ellipse in 3D of the same shape and on a frontoparallel plane or whether it comes
from a disk on a suitably slanted plane (see left below).

When there are many ellipses on the surface and if you make assumptions about the distribution
of ellipse shapes (on the surface) then an image of the many ellipses will suggest a more likely surface
slant. See the two examples below. For the example on the left below, there is a higher likelihood
that the plane is close to frontoparallel (slant near 0). For the example on the right, there is a
higher likelihood that the plane is slanted backwards at some angle. This is because, for the case
on the right, the ellipsoids near the top of the image are more foreshortened and slanted, which is
consistent with the deformation that occurs when the surface is indeed slanted back.

I sketched the likelihood of the two surfaces that have many ellipses. On the left, I sketched the
slant as having a maximum at 0, whereas on the right the likelihood of slant was some intermediate
value, say 45 deg. I emphasize that at this stage we are just talking intuitively about likelihood.
We are saying that the pattern of ellipses within each image seems more consistent with one range
of slants (and tilts) than another. To put it slightly more formally and to give you a heads up about
how likelihood is defined formally, we say that, for the pattern of ellipses in the image on the right
to arise *without* the surface being slanted, one would require a relatively unlikely set of events. For example, the 3D ellipses that project to the top of the image would just happen to be smaller than the ones that project to the bottom of the image.

Probability review

We will be talking about image random variables $I$ and scene random variables $S$. (I will assume you all know what a random variable is.) I will keep the discussion at a high level here. Let $I = i$ refer to some ‘image’, in particular, responses of cells e.g. photoreceptors, retinal ganglion cells, or simple and complex cells in V1. If we want to have any hope of doing math here, then we need to be talking about models of these cells, e.g. DOGs or Gaborś. Let $S$ be a random scene variable which could be depth or slant and tilt.

We assume that these are discrete random variables. To do so, we break the sample spaces of $I$ and $S$ into discrete chunks and consider $I = i$ to be some chunk and $S = s$ to be some chunk. For example, when we have an image, we’ve broken the infinitely many possible intensities into a finite set of possibilities, say 0 to 255.

Here are a few standard definitions that we’ll be using:

- **joint probability** $p(I = i, S = s)$
- **marginal probability**
  
  \[ p(I = i) = \sum_{s \in S} p(I = i, S = s) \]
  
  \[ p(S = s) = \sum_{i \in I} p(I = i, S = s) \]
- **conditional probability**
  
  \[ p(I = i | S = s) = \frac{p(I = i, S = s)}{p(S = s)} \]
  
  \[ p(S = s | I = i) = \frac{p(I = i, S = s)}{p(I = i)} \]
You can visualize these definitions as follows. Consider a 2D matrix where the rows are different values of $I = i$, which could represent image intensities or responses to image intensities. The columns are $S = s$ which could represent different values of a scene parameter. I have indicated vectors in the illustration and these could be image motion vectors, but the scene variables could be image orientations or surface slants or tilts. The key difference between the $I$ and $S$ is that the $I$ are measurable image quantities and the scene $S$ are quantities that must be inferred.

The conditional probability $p(I = i|S = s)$ is called the likelihood. It should not be confused with $p(S = s|I = i)$ which has a different name and which I will discuss next lecture. In the problems that I discussed earlier, I mentioned 'likelihood' intuitively and it was tempting to think it is as a probability of a scene. But that is not quite the idea. Rather the likelihood is the probability of an image, given a scene. The problem of choosing the maximum likelihood is to choose the scene $S = s$ such that the image $I = i$ arises with a greater likelihood for that scene $S = s$ than for any other scene. By 'likelihood' here, we are referring to something random, specifically the randomness of image formation. This could be noise in the imaging process, or it could be the randomness in the scene variables that generate the image, such as the sizes and shapes of the ellipses in the case of the shape from texture examples.

**Maximum likelihood for the intensity increment**

Take again the example of the intensity increment, and consider the maximum likelihood estimate for the center region. Assume that the image we measure for the center is $I(x, y) = I_0 + \Delta I + n(x, y)$ where $n(x, y)$ is Gaussian noise of mean 0 and variance $\sigma_n^2$ that is added to each pixel.

We can write the probability of one noisy pixel in the center patch of the image, given the scene variables $I_{\text{center}} = I_0 + \Delta I$ is as follows.

$$p(n(x, y) = n_i) = \frac{\Delta n}{2\pi \sigma_n} e^{-\frac{n_i^2}{2\sigma_n^2}}$$

where we have discretized the range of noise values into bins of size $\Delta n$ which are small enough that we can ignore the fact that the curve is not exactly constant within each bin. [ASIDE: I am adding the $\Delta n$ term here because someone pointed out in class that the probability of noise $n(x, y)$ taking one particular value is 0, since this random variable has a continuum of values.]
It is standard to assume the noise at different pixels is independent. (Two random variables $X$ and $Y$ are independent if their joint probability $p(X,Y)$ function is equal to the product of their two marginals $p(X)p(Y)$.) Let $N$ be the number of pixels in the center patch. Then the probability of getting $N$ noise values $n(x, y)$ for the different $(x, y)$ is:

$$p(n) = \prod_{x,y}(\frac{\Delta n}{2\pi\sigma_n})^N e^{-\frac{n(x,y)^2}{2\sigma_n^2}}$$

$$= c e^{-\sum_{x,y} \frac{n(x,y)^2}{2\sigma_n^2}}$$

$$= c e^{-\sum_{x,y} \frac{(I(x,y)-I_0-\Delta I)^2}{2\sigma_n^2}}$$

where $c$ is a constant.

The probability of a particular set of noise values $p(n)$ in the center patch is depends on the values of $I_0 + \Delta I$ and on the (noisy) image, and can be written as a likelihood:

$$p(n) = p(I \mid I_0 + \Delta I)$$

We want to find the value of $I_0 + \Delta I$ that has the highest likelihood, that is, the most probable noise.

The plot below is output from the following: http://www.cim.mcgill.ca/~langer/546/MATLAB/likelihood.m

It plots the likelihoods of $\Delta I$ when the true $\Delta I$ is 2 and there are 400 noise values. (Here we just set $I = 0$.)

The title shows the 'mean' likelihood, which I am taking as a proxy for the 'maximum' since the distribution is presumably symmetric here. (The max is harder to compute since one would need to interpolate.) The mean likelihood typically has a value close to the true value 2, but it isn’t as close as you might have expected. Run the experiment a bunch of times to see that the mean jumps around a bit around the value 2.

How can one define likelihood functions for other problems such as deciding on the orientation(s) present in the neighborhood of a pixel, or the binocular disparity or image motion or the slant and tilt of a surface? I will not go into details on this here, but see me if you are interested in working on it. It would be nice project. To motivate the problem a bit more, recall Assignment 2 where
you computed the responses of orientation detector cells (Q1) and disparity cells (Q2). For Q1, we would like to take the various responses of cells of different orientations and come up with a likelihood function for the correct orientation. (What you did in Q1 was just choose the orientation with the maximum response. But note that this is not exactly the same thing as choosing the maximum likelihood, since the responses are not identical to likelihoods. Rather, you would need to construct a likelihood, based on the responses.

Similarly, for the disparity question (Q2) you now have a good sense of the range of responses that one can get for a given true disparity and for a disparity tuned cell. The idea for a maximum likelihood estimate is to define a likelihood function based on the responses of disparity tuned cells at each point, and then select to choose the disparity with the maximum likelihood.

For shape from texture, the problem is more complex to address and I did not go into the details. Expressions for the likelihood function have been worked out by David Knill and colleagues in the 1990s using ellipses as the texture elements. These are the examples I’ve shown. The likelihood function involves two scene parameters, slant and tilt. As I’ve mentioned before, the shape parameters of the scene ellipses are randomized which induces a random distribution of ellipses in the image plane. This image ellipse distribution depends on the slant and tilt of the given scene. Knill derived the details of $p(\text{image ellipses}|\sigma, \tau)$ by treating the ellipses as independent.

**Relationship between psychometric functions and probabilities**

The above discussion was about theoretical likelihood functions that we can derive from models of the scene to image mapping and models of the noise present in such mappings. However, these models don’t necessarily predict the behavior of real observers. It is common, therefore, to fit models that are based on observer behavior, and to make prediction of other behaviors based on these fitted models.

One common idea for fitting a model of likelihoods from user’s behavior is shown below. Given a psychometric function in some experiment, fit this function using a model of blurred step edge. Formally, a blurred step edge corresponds to a cumulative Gaussian i.e. the integral of a Gaussian from negative infinity up to some value. (I just ask you to accept this fact.) The main idea is that the threshold $\Delta s$ can be chosen to correspond to some number of standard deviations (say 1) of the Gaussian that best fits the cumulative Gaussian model. Then, and this is the big step, one can take the likelihood function to be this Gaussian (multiplied by some unknown constant that we don’t care about).

[ASIDE: This is a very common modelling method. But does it make sense? What does it assume? The answer depends on the experiment that defines the psychometric curve, and how one interprets it. So let’s take an example. Often one has a discrimination task where one asks an observer if a test value is greater than a reference value. The likelihood function in this case might be interpreted as the conditional probability that the test stimulus has a particular value. To connect the psychometric function (data) to the likelihood function (model), one could interpret the probability of the observer’s response for some test stimulus level $s$ in terms of the sum of the likelihoods $p(I|S = s)$ over all $s'$ such that $s' < s$.

While this interpretation might seems a bit forced, it is difficult to come with something better. If you have a better idea, let me know. Most researchers who do this sort of work are happy with the above model. Keep in mind that a key goal of the model is to make predictions. So if the models do that, then they are useful. And these models have made many successful predictions.
about behavior. Next class we’ll discuss what sort of predictions can be made, namely when we look at ‘cue combinations’.

![Psychometric function](image1)

![Model of likelihood](image2)