

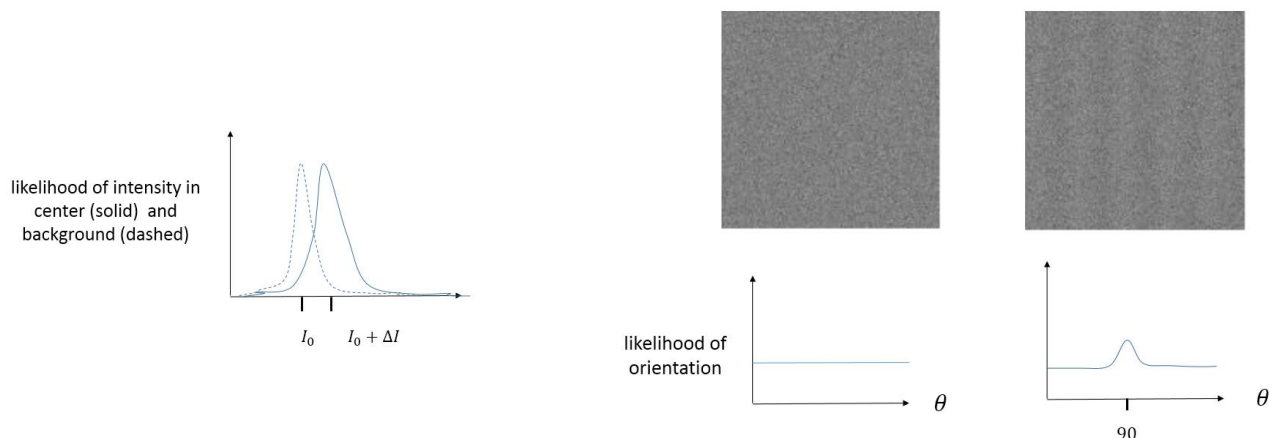
Today I will revisit some of the ideas that I introduced last lecture, and I will try to pose these ideas in terms of probabilities. I'll then use probabilities next lecture when I describe a theory of how the visual system combines different cues.

## Motivation

Recall 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. (Or pixel noise could also be added to the image intensities themselves. ) We would like to model the observer's uncertainty that comes from this noise.

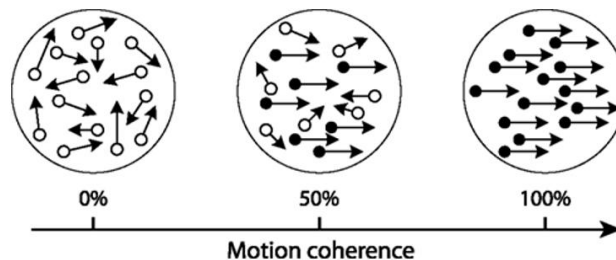
We will use the term 'likelihood' as well a probability this lecture. Likelihood has a formal definition which I will review a bit later. For now, let us take the likelihood intuitively to be proportional to the probability that a given image was the result of a background that has intensity value  $I_0$  plus noise, and that the center square has intensity value  $I_0 + \Delta I$  plus noise. So the likelihoods are a function of the uniform (pre-noise) intensity of the two image regions. These likelihood functions are sketched below on the left.

As another example, consider the orientation of a 2D sinusoid image with some additive noise. If the contrast of the sinusoid is low relative to the amount of noise, it will be difficult to discern the orientation of the sinusoid. The image will have been equally likely to occur for any orientation. As the contrast of the sinusoid is increased, the sinusoid will gradually become visible in the image and the likelihood will be elevated at the correct orientation. That is, the noisy image will be more likely to have occurred for the correct orientation than for some other orientation.

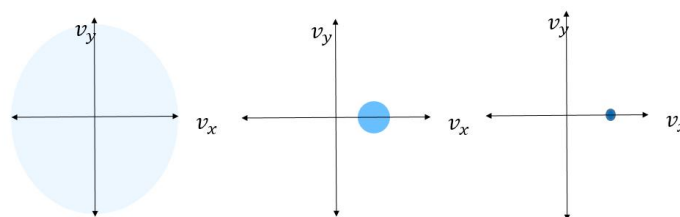


Another example is image motion. Below is a image motion stimulus which consists of dots moving in 2D. Each dot either moves with some fixed image velocity  $(v_x, v_y)$  or else it moves with some random velocity with mean 0 drawn from a distribution. The observer's task might be to judge the velocity  $(v_x, v_y)$ . We assume that the stimulus is filtered through orientation/motion sensitive cells and the percepts must be made based on the responses of these cells. To vary the difficulty of the task – that is, the level of uncertainty in  $(v_x, v_y)$  – the experimenter typically varies the fraction

of dots that move with  $(v_x, v_y)$  versus the fraction that move with the random velocity. The fraction that moves with  $(v_x, v_y)$  is sometimes called the *motion coherence*.



Below I sketch out informally the likelihood function for  $(v_x, v_y)$  for this case as the coherence increases. For 100 % coherence (right), the likelihood will be concentrated around the true velocity. There still will be some spread, however, because the stimulus consists of random dots and there are multiply possible pairings in principle from frame to frame. For 50 % coherence (middle), there will be less of a spread in the likelihood because only half the velocities are in a random direction. For 0 percent coherence (left), there will very high uncertainty in what  $(v_x, v_y)$  is. (Indeed it is not even well defined, if no dots move with that velocity.)

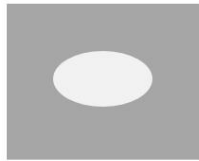


The next example is for depth from binocular disparities. (Figure omitted.) If the stimulus is a random dot stereogram with a center square protruding from a background then the likelihood function will be similar to the one on the previous page for  $I$  and  $I + \Delta I$  except that now it will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

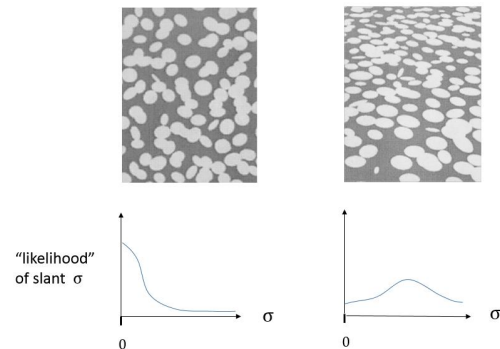
The final example for now is shape from texture. Take a class of stimuli in which the texture is generated from random shaped ellipses placed on a slanted 3D plane. The shapes and sizes and positions of the ellipses are chosen randomly from some distribution. As illustrated in the image on the left below, for a single image ellipse, it is uncertain whether it comes from an ellipse of the same shape and on a frontoparallel plane or whether it comes from a disk on a suitably slanted plane, or from other shaped ellipse on a differently slanted plane.

The texture consists of many ellipses on the surface. If the observer knows or assumes the distribution of ellipse shapes on the surface then, for any given image containing many ellipses, some surface slants will be more likely than others. 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 more dense and smaller, which is consistent with the deformation that occurs when the surface is indeed slanted back. It would be very unlikely, for example, to have a frontoparallel plane that produced that gradient in the sizes and foreshortening of ellipses observed in the right image.



Is this an ellipse on a frontoparallel plane, or a disk on a slanted plane?



## Probability review

Let's now be more formal about what we mean by "likelihood." We will be talking about image random variables  $I$  and scene random variables  $S$ . Let  $I = i$  refer to some 'image'. In practice this could refer to the 2D matrix of image intensities themselves, or it could refer to the responses of a set of cells e.g. photoreceptors, retinal ganglion cells, or simple and complex cells in V1. Or, in the case of shape from texture, it could refer to the image positions and aspect ratios and orientations of the ellipses that define the image and are assumed to be accurately measured.

Let  $S$  be a random variable that corresponds to some scene property that is manipulated in the experiment e.g. luminance, depth, orientation, binocular disparity, slant or tilt, etc. The key difference between the  $I$  and  $S$  is that the  $I$  are measurable image quantities whereas the  $S$  are scene quantities that are inferred.

We assume that these are discrete random variables. Sometimes this is already the case e.g. 8 bit images, but if not then we can partition the (continuous) sample spaces of  $I$  and/or  $S$  into bins and consider  $I = i$  to be some bin and/or  $S = s$  to be some bin. This is essentially what is already done with image intensities, namely we've broken the infinitely many possible intensities into a finite set of possibilities, say 0 to 255.

Here is the notation that we'll be using for basic probability definitions. You should be familiar with this. If not, you'll need to brush up.

- *joint probability*  $p(I = i, S = s)$
- *marginal probability*

$$p(I = i) = \sum_{s \in S} p(I = i, S = s)$$

$$p(S = s) = \sum_{i \in I} p(I = i, S = s)$$

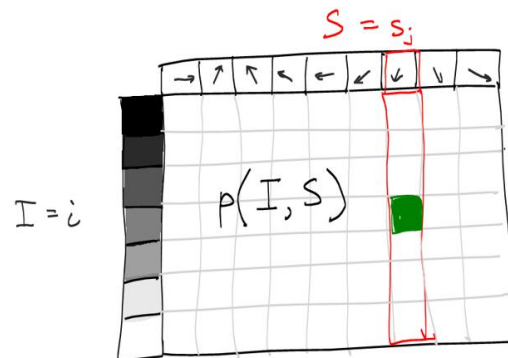
- *conditional probability*

$$p(I = i|S = s) = \frac{p(I = i, S = s)}{p(S = s)}$$

$$p(S = s|I = i) = \frac{p(I = i, S = s)}{p(I = i)}$$

You can think of the joint probability function  $p(I, S)$  definitions as follows. (See slides for illustrations of the marginal or conditional probabilities.) Consider a 2D matrix where the rows are different values of  $I = i$ . Each row is an image or the set of DOG or Gabor (or complex cell) responses to an image. The different rows are not single pixels or single responses from one DOG or Gabor cell, etc, but rather they are entire images or sets of responses.

Each column  $S = s$  could represent different values of a scene parameter. I have indicated vectors in the illustration: these could be 2D image motion vectors, or image orientations (a 2D sinusoid) or surface slants or tilts.



## Likelihood function

The *likelihood* function is the conditional probability  $p(I = i|S = s)$ . It depends on both  $I$  and  $S$ . 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 that an image  $I = i$  occurs, in the case that the scene was  $S = s$ . If we fix  $s$  and vary  $i$ , then  $p(I = i|S = s)$  is a probability density function over  $i$  and it integrates to 1. However, that’s not what we are interested in here. Rather, the image  $i$  is given and we are comparing different scenes  $s$  as possible ways of explaining  $i$ . If we integrate over  $s$  (for fixed  $i$ ), we do not get 1. That is why we do not call  $p(I = i|S = s)$  a probability of  $s$ , and we instead some other word was invented, namely ‘likelihood’.

The vision system doesn’t know what this scene  $s$  is. It is only given the image  $I = i$ . The *maximum likelihood* method is to choose the scene  $S = s$  that maximizes  $p(I = i|S = s)$ , that is, 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, namely the randomness of image formation. We are not referring to the randomness in  $S$ , since  $S = s$  has already occurred. A few examples should help with this rather subtle distinction.

## Maximum likelihood for an intensity increment

Take again the example of the intensity increment. Suppose we have a background region of some intensity  $I_0$  and we have a center square region of some intensity  $I_0 + \Delta I$ . Assume that the visual system knows the location and size of the square. Also assume that noise is added to each pixel in center and surround. For example, one often assumes that the noise values  $n(x, y)$  have a Gaussian probability density with mean 0 and variance  $\sigma_n^2$ . As mentioned earlier, image intensities are quantized into discrete values, so we would need a discrete approximation but let's not concern ourselves with that detail.

Suppose the task is to decide if there an intensity increment as opposed to decrement. Say one solves this by estimating  $I_0$  in the surround and  $I_0 + \Delta I$  in the center. Let's just take the case of estimating  $I_0$  in the surround. We can write the probability of one noisy pixel in the surround, given  $I_0$  is as follows.

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

The constant  $c$  is there because the Gaussian is continuous density and I want to write the probability as discrete, i.e. we discretize the range of noise values into bins.

The probability of a particular set of noise values  $p(n)$  in the surround patch depends on the values of  $I_0$  and on the (noisy) image  $i = I_0 + n$ , and can be written as a likelihood:

$$p(n) = p(I = i \mid I_0)$$

We want to find the value of  $I_0$  that has the highest likelihood, that is, the  $I_0$  value such that the noise  $n$  required to produce the given image  $i$  would have had the highest probability of occurring.

It is standard to assume the noise at different pixels is independent.<sup>1</sup> Let  $N$  be the number of pixels in the surround region. Then the joint probability of the  $N$  noise values  $n(x, y)$  for the different  $(x, y)$  is the product of probabilities of noise for the individual pixels:

$$\begin{aligned} p(n) &= \prod_{x,y} \left( \frac{c}{\sqrt{2\pi}\sigma_n} \right)^N e^{-\frac{n(x,y)^2}{2\sigma_n^2}} \\ &= \text{constant} * e^{-\sum_{x,y} \frac{n(x,y)^2}{2\sigma_n^2}} \\ &= \text{constant} * e^{-\sum_{x,y} \frac{(I(x,y) - I_0)^2}{2\sigma_n^2}} \end{aligned}$$

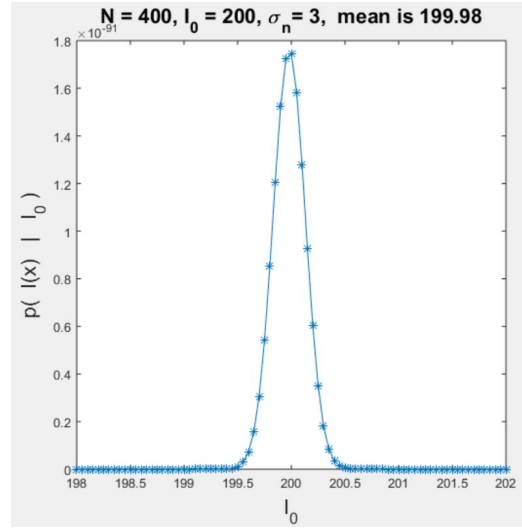
If you want to see a particular example, then here is some matlab code.

<http://www.cim.mcgill.ca/~langer/546/MATLAB/likelihood.m>

The plot below was generated from that code. It plots the likelihoods of  $\Delta I$  when the true  $\Delta I$  is 200 and there are 400 noise values. (Here we just set  $I = 0$  for simplicity.) If you run the code a few times, you'll see that the likelihood function changes quite a lot between runs. Even with 400 samples, there is a lot of variability.

Note that the title shows the 'mean' likelihood, which I am taking as a proxy for the 'maximum' since the distribution is roughly symmetric here. (The max is harder to compute since one would need to interpolate between the discrete samples.) As you can see if you run it a few times, the mean likelihood typically has a value close to the true value of  $I_0 = 200$ .

<sup>1</sup>Two random variables  $X_1$  and  $X_2$  are independent if for any  $X_1 = x_1$  and  $X_2 = x_2$ , their joint probability  $p(X_1 = x_1, X_2 = x_2)$  function is equal to the product of their marginal probabilities  $p(X_1 = x_1)p(X_2 = x_2)$ .



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 the mathematics here because it is more an exercise in probability than in vision. But let's at least sketch the idea of how you could construct such a model.

Take the case of binocular disparity. Recall Assignment 2 where you computed the responses of V1 complex cells tuned to particular disparities, or you considered MT cells tuned to particular image velocities  $(v_x, v_y)$ . In each case, it is possible in principle to write down a mathematical model. For example, take a region of constant disparity  $d$ , the likelihood of the responses  $r_{d_1}, r_{d_2}, \dots, r_{d_k}$  of cells tuned to different disparities, if the actual disparity was  $d$ . One could try to write down a model:

$$p((r_{d_1}, r_{d_2}, \dots, r_{d_k}) \mid \text{disparity} = d).$$

This is not easy to do, but it can be done. You could something similar for the 2D motion estimation problem. You have shift detector cells tuned to different orientations and motions. You could come up with a likelihood function for the responses of these cells:

$$p((r_{\theta_1, \text{speed}_1}, \dots, r_{\theta_i, \text{speed}_j}, \dots) \mid \text{actual velocity is } (v_x, v_y)).$$

where  $r_{\theta_i, \text{speed}_j}$  is the response of cells with peak tuning to spatial orientation  $\theta$  and  $\text{speed}_j$  in that orientation. Again, not easy, but it can be done.

For shape from texture, it is also possible to come up with likelihood functions. I mentioned the work of David Knill whose model assumed that textures were ellipses distributed over a slanted plane. The ellipses on the surface had a random distribution, in size, orientation, and elongation (aspect ratio) and Knill wrote out precise mathematical model for this. He also considered the projection of the ellipses into the image, which gave rise to image distributions of size, orientation, and elongation. He was able to write down likelihood functions of the form:

$$p(\text{image ellipses} \mid \text{surface slant})$$

This allowed him to estimate the maximum likelihood of a surface slant for a given image. Note that there is no intensity pixel noise here. Rather, the randomness is in the distribution of ellipses themselves and it is assumed that the ellipses can be measured in the image.

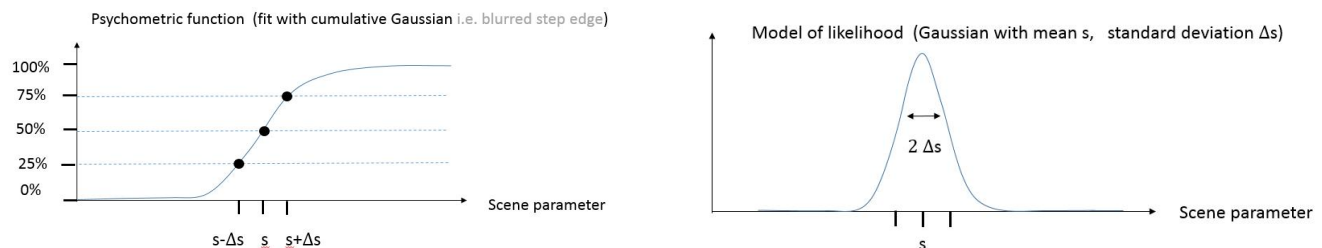
## Likelihood functions and psychometric functions

The mathematical models that I mentioned above don't necessarily predict the behavior of real observers. How can we relate real observer behavior to such models?

One common approach is illustrated below. Given a psychometric function in some experiment, one fits this function using a cumulative Gaussian function which by definition is the integral of a Gaussian from negative infinity up to some value  $x$ :

$$cdf(x) \equiv \int_{-\infty}^x G(x', \mu, \sigma) dx'$$

where  $cdf$  stands for “cumulative density function”. Here  $\mu$  is the mean and  $\sigma$  is the standard deviation for Gaussian.



The experiment might be to say whether some test stimulus has a greater or smaller scene parameter  $s$  than some reference stimulus. We've seen several examples:

- a background intensity  $I$  and a central square with intensity  $I + \Delta I$
- a background random dot pattern with disparity  $d$  and a central square with disparity  $d + \Delta d$

The subject's performance in the task is described by the psychometric function. If one fits the performance using a cumulative Gaussian model with mean at some scene parameter value  $s$  (e.g. the  $I$  or  $d$  in the above examples), then the subject's uncertainty in doing the task can be associated with the standard deviation of the Gaussian whose cdf is the best fit to the psychometric function.

It is common to treat this fitted Gaussian as if it were the person's likelihood function<sup>2</sup> which they use to estimate the scene parameter  $s$ . As we will see next lecture, this is useful for considering how people combine different visual cues.

<sup>2</sup>We don't really believe that people have a likelihood function in their brains — any more than we believe that your brain solves differential equations when you walk and throw a ball. But let's not get into that philosophical issue here.