Computational Vision
U. Minn. Psy 5036
Daniel Kersten
Lecture 3: The Ideal Observer

Goals

Last time

Limits to light discrimination

In the famous experiment of Hecht, Schlaer, and Pirenne, measurements were made of the proportion of times observers saw a dim small briefly flashed spot of light. Under the right conditions, a person can detect an amazingly small number of photons. Perhaps even more remarkable was the discovery that single photoreceptors are capable of transducing single photons. We singled out the problem of variability in the photon emissions and absorptions as a key computational problem to understand.

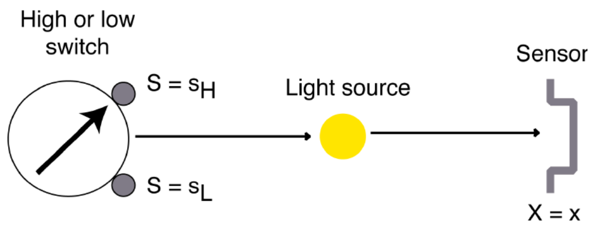
Why was the subject response curve sigmoidal? If there was a specific light intensity (or number of photons) at transition from invisible to visible, we should see a "step" function--but we don't. The answer is *variability or "noise".* But where does it come from? The variability may be due to several causes: the statistics of photon emission and absorption we've mentioned, "noisiness" in human responses, or in neural transmission. The physical limits to visual detection and discrimination can be traced to the stochastic nature of photon absorption and emission. This stochastic nature or variability is a consequence of the particle nature of light. A careful understanding of these limits in conjunction with psychophysical measurements is an essential component to understanding human sensitivity.

Today

We study key concepts in Signal Detection Theory (SDT) and ideal observer analysis in the context of light intensity discrimination. The essence of the ideal observer approach is to ask: What is the optimal or ideal strategy for a task? What are the theoretical limits to performance on a task? In particular for our example, what are the limitations on the discrimination of light levels that any observer must face?

Visual decisions as inference, hypothesis testing

Inference is choosing an hypothesis based on data. In this course, we will treat hypotheses as variables, so we can be more precise by defining inferences as deciding on the value of an hypothesis (variable) given some data. In statistical inference, the decision is based on a model of the probabilities of the hypotheses, H and the data, x. What are our hypotheses? In a discrimination experiment, we can define H = S = $\{S_L, S_H\}$. In a detection experiment: H = S = $\{S_{\text{present}}, S_{\text{not present}}\}$. (Recall that Hecht et al. didn't measure responses under the condition: S = $S_{\text{not present}}$. We now appreciate the importance of measuring responses under both conditions.)



In[1]:=

We can also think of the visual detection or discrimination task in terms of signal transmission. The "sender" wishes to set the light switch to one of two positions, corresponding to a high and low average light intensities (or on and off). The signal space corresponding to the states of the switches: H = S_L (low), or H = S_H (high). Sometimes it is also useful to think of the hypothesis as a cause of the data.

As we will see, the generative (or transmission) problem is completely characterized by: 1) the probability, p(H), which in our simplified case is just two numbers corresponding to the probability p(H=S_L) and p(H=S_H); 2) the probability p(x | H), which is characterized by two (infinitely long) lists of numbers, a list for the probability distribution for each of the two hypothesis values.

Inference inverts the generative problem. Inference is guessing a value of H given a measurement, e.g. of x photons. We've noted that because of variability in the generation of data, there is an inherent ambiguity in determining the value of H, is it S_L or S_H? In this and the next lecture we will see how to make the best decision, given a precise definition of the generative model. Initially, we will define "best" as the strategy that leads to the fewest errors on average. A theoretical observer that makes the best decisions is called an *ideal observer*. The theory behind the ideal observer is called signal detection theory (SDT).

The computational theory of signal detection and estimation was developed in the 1940's and 1950's. More information about **Signal Detection Theory (SDT)** can be found in the classic book by Green, D. M., & Swets, J. A. (1974). Treatments from an engineering perspective can be found in Van Trees, H. L. (1968) and Poor (1994).

Ideal observer analysis

In terms of scene (S) and images (I), our approach can be summarized as:

1) model how images are generated from states of the world, or scenes: S -> I
2) model how to make the best guess, or inference of a state of the world, or a scene given a single image: I -> S'

Today, we will:

- get an intuitive introduction to the basics of probability distributions, cumulative distributions.
- get a simple introduction to Bayesian inference
- formally develop an ideal observer that models "external variability" inherent to the generative process.
- Introduce the yes/no task, and define the two ways of being correct, and the two ways of being wrong

In the next lecture, we will show how use to use these models to better understand the limits to the human ability to detect image signals. In general, a human observer introduces additional variability that may not be included in the generative model. For example, while you might be able to count every single dot in the demo in the last lecture (at least for small numbers of dots), you can't count individual photons, but you can judge brightness. So we'd guess that human performance will be sub-optimal, but by how much? Do human brightness perception behave *as if* it is missing most of the photon counts? This question will lead us to develop "ideal observer analysis" in which we:

3) compare human and ideal performance. Call the guesses by the human observer S''. Ideal observer analysis compares S'' to S' for a criterion level of performance, e.g. where human and ideal get the same proportion correct.

Side note: Some conceptually related terms: (causes, scene properties, signals, hypotheses, states of the world) -> (data, image intensities, image measurements, features).

Motivation: a computational theory of discrimination

Let's consider the light discrimination task--where you will be the observer that has to decide which of two images has more dots, analogous to deciding which of two images is brighter, i.e. has more photons.

Model for light, two switch settings

As above, let's define a Poisson distribution with a mean represented by the variable name **mean**, with a function to draw a sample from this distribution. Let highmean = 7 for the "high" setting, and lowmean =5 for the "low" setting. So at a given time, **mean** can take on only one of two values: **mean** = highmean or lowmean.

In[119]:=
```
numberofphotons[mean_] := RandomInteger[PoissonDistribution[mean]];
dotsize = 0.01;
highmean = 7;
lowmean = 4;
```

We'll define a "blank" graphics display that we'll use to "turn off" the display--i.e. make it black.

In[123]:=
```
blank = Graphics[{PointSize[dotsize], Black, Point /@ {{}, {}}},
   AspectRatio → 1, Frame → False, FrameTicks → None,
   Background → GrayLevel[0.0], PlotRange → {{-0.2, 1.2}, {-0.2, 1.2}}];
flash =
  blank;
```

We could keep the display within this notebook, but let's create a new one and insert the graphics object "flash" in it, and at the same time declare it as a **Dynamic** variable. We do this so that it will automatically change whenever we do something to the flash variable in our home notebook.

In[125]:=
```
CreateDocument[Dynamic[flash], ShowCellBracket → False, WindowSize → {300, 300},
  WindowElements → {}, Background → Black, NotebookFileName → "Flash Display"];
```

Now we'll create a function called "twoflashes" that we can call whenever we want to display a pair of flashes in our "Flash Display" window. We use **RandomInteger[1]** to simulate a "coin flip" to decide whether the switch (our "state of the world" or signal) is set to high or low.

In[126]:=
```
twoflashes := Module[{tempmean},
  If[RandomInteger[1] == 1, firstmean = lowmean;
   secondmean = highmean, firstmean = highmean;
   secondmean = lowmean];

  numhighsample = numberofphotons[firstmean];
  highsample = Table[RandomReal[{0, 1}, 2], {numhighsample}];
  (*pick a random 2D location for the dot*)
  flash = Graphics[{PointSize[dotsize], Red, Point /@ highsample},
    AspectRatio → 1, Frame → False, FrameTicks → None,
    Background → GrayLevel[0.0], PlotRange → {{-0.2, 1.2}, {-0.2, 1.2}}];
  Pause[1]; flash = blank; Pause[.5];
  numhighsample = numberofphotons[secondmean];
  highsample = Table[RandomReal[{0, 1}, 2], {numhighsample}];
  flash = Graphics[{PointSize[dotsize], Red, Point /@ highsample},
    AspectRatio → 1, Frame → False, FrameTicks → None,
    Background → GrayLevel[0.0], PlotRange → {{-0.2, 1.2}, {-0.2, 1.2}}];
  Pause[
   1];
  flash = blank;]
```

Whenever you evaluate the following cell, it should present a pair of "flashes" in the "Flash Display" notebook.

In[133]:=
```
twoflashes
```

Now you be the "sensor" that counts dots (aka “photons”)--and with each flash, decide whether it was the first or the second flash that was caused by the high switch setting.

This is an example of a *two-alternative forced-choice experiment* (2AFC). More on this later.

What do think is the best strategy that enables you to get the highest proportion of correct guesses? Intuitively, it is just what you’ve been doing--count the number of dots in each and pick the image with the most dots. And you’d be right. Despite the obviousness of the intuition, today we are going to prove it. The rationale is that by understanding how to prove it, you’ll understand how one can derive optimal decision rules for more complex problems, where the intuition isn’t as obvious.

Preview of needed concepts & tools

Probability concepts

In the next lecture, we will systematically go over the elements of probability and inference in detail. But today, we will take an intuitive approach, letting the light discrimination problem motivate the need for the concepts we need. We will motivate the later use and development of the concepts of: **Random variable**, **Probability distribution**, **Conditional probability**, **Prior probability**, **Posterior probability**.

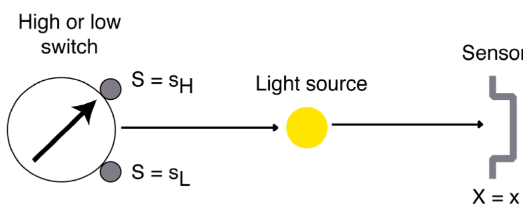
Below we are going to use the following built-in functions:

Table, Range, Map, Count, ArrayPlot, Show, PoissonDistribution, PDF, RandomVariate. We will also use the “pure” function, Function[ ], but in short-hand form.

Modeling external variability: The Ideal Observer

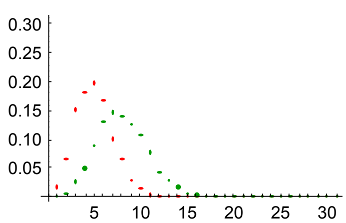
Generative model for stimulus: *Given the switch setting, what measurements result?*

Our first goal is a quantitative description of distribution of photons measured under the two switch settings. We will simulate the process on a trial-by-trial basis, and then compare the simulation with theory.



Histograms and population distributions

Imagine first that we do an experiment with a setup as above. First set the light level to S_L. Now we begin to send flashes, e.g. by opening and closing a shutter. As we’ve emphasized, no matter how hard we try to make the apparatus perfect, we would discover that the photon counter doesn't always record the same number of photons for each flash. Then we do the same for the switch setting, S_H. So, we compile a histogram of photon counts and find something like this:



The horizontal axes represents a photon count, and the vertical axis represents the proportion of times that a particular count was measured. The histograms provide a quantitative representation of the variability. As the number of samples grows (i.e. test flashes in our case) , the histogram becomes a better and better approximation of a “true”, underlying Poisson distribution. Although we don’t prove it here, the Poisson distribution arises because of the independence of photon absorption.

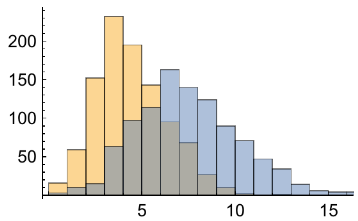
Simulation of the generative process-- Photon counts, and histograms corresponding to high and low switch settings

Let’s compile two histograms of photon counts conditional on the switch settings. Define high and low mean models:

In[38]:=
```
nsamples=1000;
zlow = RandomVariate[PoissonDistribution[4],nsamples];
zhigh = RandomVariate[PoissonDistribution[7],nsamples];
```

How frequently do trials with 1, 2, ...n dots occur? As at the end of the previous lecture, we use the built-in function: **Histogram[]** to show us,
using: **Histogram[{zlow,zhigh}]** .

In[41]:=
```
Histogram[{zlow, zhigh}, ImageSize → Small]
```

Out[41]=


Compare theoretical Poisson probability distributions to high and low simulated histograms

In[14]:=
```
PDF[PoissonDistribution[μ], x]
```

Out[14]=
$$\begin{cases} \frac{e^{-\mu}\mu^x}{x!} & x \geq 0 \\ 0 & \text{True} \end{cases}$$

In standard probability notation, if the mean value is μ, then the probability of random variable X taking on the value of x photons is:

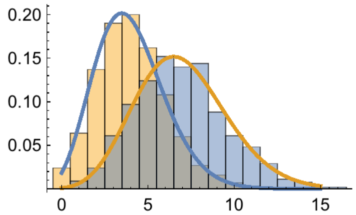
$$p(X=x \text{ photons}) = \frac{\mu^x e^{-\mu}}{x!}$$

The Poisson distribution is characterized by the mean (or expected) number of photons flashed when the switch is set to S_L. The spread in the distribution (measured by the standard deviation) is equal to the square root of the mean. The mean or expectation of X is $E(X) = \mu$, and for a Poisson distribution, the variance, $var(X) = E[(X - \mu)^2]$ is also equal to μ.

▸ 1. Evaluate: Variance[PoissonDistribution[μ]], Mean[PoissonDistribution[μ]], Kurtosis[PoissonDistribution[μ]], CDF[PoissonDistribution[μ],x]

We will use the theoretical expression for the Poisson distribution below. But here, let's make a quick comparison of the sampled histogram with theory:

In[42]:=
```
Show[Histogram[Table[RandomVariate[PoissonDistribution[i], nsamples], {i, {4, 7}}],
   {1}, "PDF", PlotRange → All],
 Plot[Evaluate[Table[PDF[PoissonDistribution[μ], x], {μ, {4, 7}}]],
   {x, 0, 15}, PlotStyle → Thick, PlotRange → All], ImageSize → Small]
```

Out[42]=


▶ 2. Play with the **nsamples** variable. Note that the more samples we draw, the better the experimental histograms match the theoretical. Try a small value of nsamples, say **nsamples = 100**;

Conditional probabilities: a preview

Our histograms are estimates of the *conditional probabilities* p(x|S_L) and p(x|S_H). I.e. they represent the probability of observing a photon count of x, given (or "conditional on") the switch setting H = {S_L or S_H}. We will treat the switch setting as an hypothesis, which is also a random variable, i.e. H. As we've noted earlier, the hypothesis is about the state of the world, such as a signal state (high or low setting), or an object property.

A note on terminology: p(x|S_L) is the probability of x conditional on S_L. But below when we consider optimal inference, we will talk about likelihood functions of H, p(x|H), where for example p(x|$H = S_L$) is the *likelihood* of S_L given a measurement of x. A likelihood function of H is not a probability distribution over H.

The knowledge from a generative model gets embodied in the conditional probability--given a cause H (where H = S_L or S_H), the conditional probability tells us the probability of a measurement x. In a concise form, p(x|H), is one component of the generative model, and as we see below, p(H) (the "prior") is the other part.

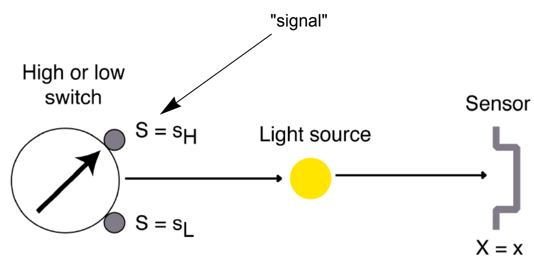
Inference model for discrimination: *Given a measurement, what was the switch setting?*

Our goal now is to show how the use of the photon (or dot) count arises from a deeper principle: The best guess should be based on choosing the most probable switch setting given a measurement. This is called choosing the setting that "maximizes the posterior probability".

Yes/no task

In a yes/no task, we designate one switch position as the "signal" and the other as "not the signal", or "noise". We'll have the observer say "yes" when she/he guesses that the switch is set to high.

In[16]:=



Posterior probability distribution

Given a measurement, what was the switch setting? We've seen that because of variability in the measurement, one can't answer this question correctly every time just from the data. However, we can ask how the decision should be made so that an observer is correct as often as possible (equivalent to minimizing the probability of error, or equivalently maximizing the probability of being correct). So we ask:

What decision strategy will minimize the probability of error?

The observer that achieves a pre-defined optimality criterion is our *"ideal observer"*. We'll see other ideal criteria later.

Let's go ahead and and assume that the ideal observer should decide pick the hypothesis value (signal or not) which has the bigger probability of the two. In other words, measure **x**, and then calculate:

If P(S_L| x) > P(S_H|x), then choose S_L
otherwise, choose S_H.

P(S_L | x) is "the probability of the switch taking on the value of S_L, conditional on a measurement x", and similarly for P(S_H | x).

This rule is called the "maximum *a posteriori*" or MAP rule. Note that there probabilities aren't the conditional probabilities we calculated and graphed above!

P(S_L | x) and P(S_H | x) are called *posterior* probabilities. Why "posterior"? Even without any data measurements, knowledge of P(S_L) and P(S_H) can provide important information for optimal decisions, and these are called *prior* probabilities, i.e. the information the observer has *prior* to the arrival of the data measurements. Posterior probabilities are "informed by the data".

It is often hard to directly specify the posterior probability, and it can be easier to model the prior probabilities and the likelihoods, and then use Bayes rule to relate them to each other:

$$p(H \mid x) = \frac{p(x \mid H)\, p(H)}{p(x)} \propto p(x \mid S)\, p(S) \qquad (1)$$

..ah, looks like now we will be able to plug in the conditional probabilities we calculated earlier. What we have done is to partition the generative components into two parts: a likelihood, p(x|H) and a prior, p(H).

We'll see later why Bayes' rule is true. Bayes' rule is a classic rule in probability theory, attributed to the Rev. Thomas Bayes (1702-1761). However, while Bayes introduced the idea of inverse probabilistic inference, it was later, in the early part of the 18th century, that Laplace produced the above equation. After you've gone over the basics of probability (see ProbabilityOverview.nb) you'll be able to easily prove it.

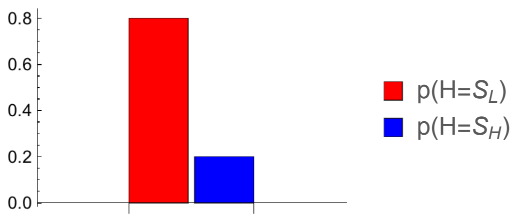
Optimal performance is determined by the posterior probability, and that is why Bayes' rule is of central importance in determining optimal theories of inference. Let's consider the roles the priors and likelihoods play separately.

Prior only

Suppose the switch is set to S_L 4/5 of the time, and to S_H 1/5 of the time.

In[164]:=
```
pH = {4 / 5, 1 / 5};
(*pH={1/2,1/2};*)
BarChart[pH, ChartStyle → {Red, Blue},
 ChartLegends → {"p(H=S_L)", "p(H=S_H)"}, ImageSize → Small]
```

Out[165]=



But if the photon (or dot) counter isn't working at all--there are no data. What then is an observer's optimal strategy to guess whether it was the high or low switch?

If the receiver's goal is to minimize the average probability of error, then it should always pick $\boldsymbol{S_L}$. Or to use the signal convention, always decide "signal not present".

Like the maximum a posteriori rule, the rule that minimizes the error rate is:

choose S_H if p (S_H) > p(S_L)
choose S_L if p(S_L) > p(S_H)

Interesting enough, people often don't do follow the optimal rule in many decision tasks. Imagine you are the observer--wouldn't it be really hard to say "no signal" every single time? With feedback, human observers and animals tend to *match* probabilities--i.e. respond S_H and S_L with the same frequencies as p (S_H) and p(S_L), respectively. (In our example, evidence for optimal decisions would be if the frequency of responding S_L = 100%). Probability matching is a behavioral phenomenon which has been studied by cognitive psychologists and economists, cf. Green et al., 2010.

Following the above optimal rule, S_H is presented p(S_H) fraction of the time--so p(S_H) = 1/5 is exactly the error rate, i.e. the fraction of the time that the observer gets the wrong answer using MAP, the minimum error rule.

Here is a simple simulation of the process, and the error rate after **numtrials**:

In[166]:=
```
numtrials = 100000;
trials = RandomVariate[BernoulliDistribution[1 / 5], numtrials];
(*Bernoulli sampling is just a coin flip model--
 biased in this example to favor 0 over 1 in our case*)
idealdecisions = Table[0, numtrials];
(*0 stands for always deciding "no signal"*)
Count[MapThread[Equal[#1, #2] &, {trials, idealdecisions}], False] / numtrials // N
Rationalize[%, .01]
```

Out[169]= 0.20005

Out[170]= $\frac{1}{5}$

Of course we could just calculate: **Count[trials,0]/numtrials**. But the above code is more general, as applied here:

▶ 3. Exercise: what is the probability of error if the probabilities are matched? Here is a simulation of the process illustrating that the error rate is higher than ideal. But can you derive the answer?

In[171]:=
```
probmatchdecisions = RandomVariate[BernoulliDistribution[1 / 5], numtrials];
(*randomly decide "0" 4/5th of the time*)

Count[MapThread[Equal[#1, #2] &, {trials, probmatchdecisions}], False] / numtrials //
 N(*count errors*)
```

Out[172]= 0.32048

Likelihood

Let's again use Mathematica's built-in Poisson function to define **p(x|H) = poisson[x, a]**, where the hypothesis or switch setting H, determines the average number of samples, x:

In[173]:=
```
poisson[x_, a_] := PDF[PoissonDistribution[a], x]
```

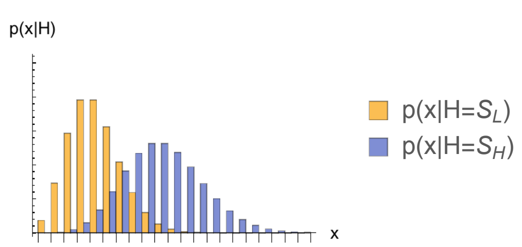
Then make two lists of probabilities corresponding to means of 4 and 9:

In[174]:=
```
p1 = Table[N[poisson[x, 4]], {x, 0, 20}];
p2 = Table[N[poisson[x, 9]], {x, 0, 20}];
```

Check that we got it right, using **BarChart[]** to view the distributions.

In[176]:=
```
BarChart[Transpose[{p1, p2}], PlotRange → {0, 0.25}, AxesLabel → {"x", "p(x|H)"},
 ChartLegends → {"p(x|H=S_L)", "p(x|H=S_H)"}, ImageSize → Small]
```

Out[176]=



Same thing we saw earlier, the plot shows the frequencies of observing x dots under the two possible switch settings, H = S_L and H = S_H.

Let's put the two likelihoods together into one 2x21 array:

In[177]:=
```
poispxH = {p1, p2};
Dimensions[poispxH]
```

Out[178]=
```
{2, 21}
```

By looking at Bayes rule (equation 1, above), you should be able to see that choosing the hypothesis with the largest likelihood is the same as choosing the hypothesis that maximizes the posterior probability when prior probabilities are both the same. When the prior probabilities equal a constant (in this case 1/2), the prior distribution is said to be *uniform*.

So either strategy will minimize the probability of error. But what if the priors are not uniform?

Combining the prior and likelihood: Maximum a posteriori (MAP) rule

By definition, the ideal "knows" the likelihoods: $\mathbf{p(x|S_L)}$ and $\mathbf{p(x|S_H)}$ and the priors, $\mathbf{p(S_L)}$ **and** $\mathbf{p(S_L)}$**.** Bayes' rule tells us how to combine them to obtain the *a posteriori* probability of the hypotheses conditional on the data x:

$$p(H \mid x) = \frac{p(x \mid H)\, p(H)}{p(x)} = \frac{p(x \mid H)\, p(H)}{\sum_H p(x \mid H)\, p(H)}$$

(The expression for the denominator p(x) follows from the basic laws of probability, see notebook / pdf.).

Let's first assume uniform priors:

In[179]:=
```
pH = {1 / 2, 1 / 2};
```

In[180]:=
```
pH = {4 / 5, 1 / 5};
```

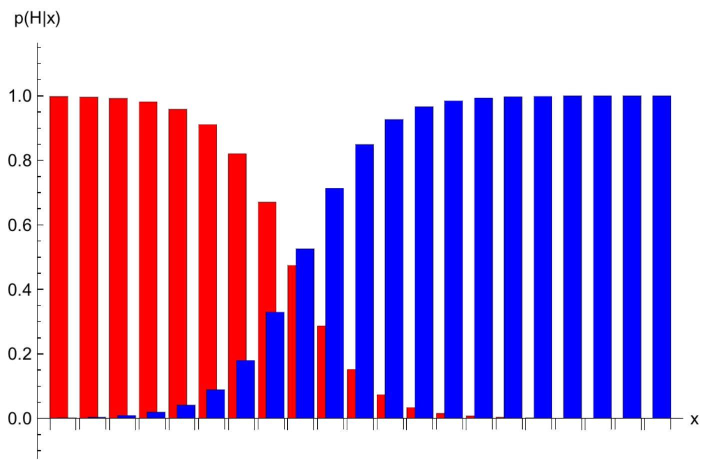
For our example, we can calculate the posterior using the following, somewhat opaque, line of *Mathematica* code:

In[181]:=
```
posterior = Transpose[Transpose[(poispxH pH)] / Plus @@ (poispxH pH)];
```

Below is a plot of the posterior probabilities, $\mathbf{p(S_H|x)}$ and $\mathbf{p(S_L|x)}$, as function of the data **x**:

In[182]:=
```
BarChart[Transpose[{posterior[[1]], posterior[[2]]}],
 PlotRange → {-.1, 1.1}, AxesLabel → {"x", "p(H|x)"},
 ChartStyle → {Red, Blue}, BarSpacing → -.55, ImageSize → Medium]
```

Out[182]=



Note that the plot shows the probability of the switch being low or high, i.e. p(H=S_L | x) or p(H=S_H|x) given that x photons were measured, or x dots were counted. (Why does the posterior plot look so different from the likelihood plot?)

Let's apply the *maximum a posteriori* rule (MAP)

For *maximum a posteriori* estimation, the rule is:

choose S_H if p(S_H|x) > p(S_L|x)
choose S_L if p(S_L|x) > p(S_H|x)

More generally, a function which chooses the value of argument (H = S_L or S_H) that maximizes a function is sometimes written as:

$$\underset{H}{\operatorname{argmax}}\, p(H|x)$$

► 4. Suppose 7 dots were counted, pick the most probable switch setting by inspecting the above graph. *NOTE: the position of the first bar corresponds to 0 photons/dots.*

► 5. Make a new plot with pH = {4/5, 1/5}. Now what is the most probable hypothesis for x=7?

Let's apply the Maximum likelihood rule

In maximum likelihood estimation, the rule is:

choose S_H if p(x | S_H) > p(x | S_L)
choose S_L if p(x | S_L) > p(x | S_H)

Because of Bayes rule, if the prior probabilities are equal, then MAP is equivalent to maximum likelihood.
p(x | S_H)/p(x | S_L) is called the likelihood ratio.

► 6. Suppose 7 dots were counted, what is the maximum likelihood decision? (See p(x|H) plots above)

Simple and classic OPTIMAL Rule for discrimination: Compare the count to a fixed criterion.

Let's assume the prior probabilities are equal, so we base our decisions using the maximum likelihood rule. We will use *Mathematica* to do some symbolic manipulations to find the transition point in our data measurements (i.e. dot or photon count) where the likelihood goes from favoring a low to a high switch setting. This is where the likelihood ratio is 1, or the log of the ratio is zero.

Let d and b be the mean number of dots for the low and high distributions, respectively.

In[226]:=
```
Clear[b, d, x];
loglikelihoodratio = Log[poisson[x, b] / poisson[x, d]];
FullSimplify[%, x ≥ 0]
```

Out[228]= $\text{Log}\left[b^x d^{-x} e^{-b+d}\right]$

In[229]:=
```
PowerExpand[%]
```

Out[229]= $-b + d + x\,\text{Log}[b] - x\,\text{Log}[d]$

There is a simple point to make here--the dot count, x, is monotonically related to the log likelihood. This means that our intuition that we can't do any better than simply counting dots was correct. But we need to decide on an appropriate criterion.

In[222]:= $\text{Solve}\left[\text{Log}\left[\frac{b^x e^{-b}}{x!}\right] == \text{Log}\left[\frac{d^x e^{-d}}{x!}\right], x\right]$

Out[222]= $\left\{\left\{x \to \frac{b-d}{\text{Log}[b]-\text{Log}[d]}\right\}\right\}$

If d and b are the average number of dots for the low and high switch settings, and the dot counts are distributed according to a Poisson distribution, then minimal error will result on average with the following rule:

Say "high" if

$$x > \frac{b-d}{\text{Log}[b]-\text{Log}[d]}$$

Say "low" otherwise.

The dot count, x, is said to be the *decision variable*. Because it is monotonically related to the likelihood ratio, it is an optimal decision variable. The transition value is called the decision *criterion*:

$$\text{decision criterion} = \frac{b-d}{\text{Log}[b]-\text{Log}[d]}$$

So now you have all you need to know to write a program that would make optimal decisions about high vs. low--it would simply count dots, and base its decision on whether the count was greater or less than the decision criterion.

Well we had guessed that counting dots was the thing to do. Now we've proved it, and now we can also calculate the criterion given the mean values of the two Poisson distributions.

Summarizing performance using performance metrics

So far so good. But we would like to go a littler farther so that we can calculate theoretically what the average performance of an ideal observer would be.

In a yes/no task, an ideal observer can make two kinds of mistakes. It can decide the signal was there when it wasn't (a **false alarm** also called a false positive), or it can decide it wasn't there when it was (a **miss** also called a false negative). These two determine the error rate.

An ideal can also be correct in two ways. It can have a **hit** (true positive) or a **correct rejection** (true negative). We'll see below that you only need to measure two of these four statistics, because the hit + miss rate have to add up to 100% (if the signal is presented 100 times, you either say yes or no, so the sum of the hit and misses has to be 100).

We develop summary measures of discrimination performance in the next lecture.

Next time: A linear, gaussian ideal observer

In the next lecture, we are going to study hit and false alarm statistics using a linear generative model in which the image measurements are the sum of a signal and gaussian noise. This formulation separates the signal from the noise, by putting the variability in the additive noise term. It has the advantage of theoretical convenience and is more easily extended to other types of signal discrimination and detection problems.

We've see today how to model the variability in the generative process and how to compute the optimal or ideal observer. Next time we'll look more closely at modeling internal variability of non-optimal observers such as ourselves.

We'll see how to calculate the signal-to-noise ratio (d') for a discrimination or detection task in terms of hit and false alarm rates, and other measures, for both ideal and human observers.

You've probably noted above that when you are a subject in a yes/no task it is hard to decide and maintain your personal criterion. Feedback can help a human observer learn to at least maintain consistency, but we'll see that a better way of measuring performance is to use the two-alternative forced-choice (2AFC) task.

Ideal observer analysis will allow to benchmark human performance in terms of a single number, *statistical efficiency* relative to ideal performance.

References

Applebaum, D. (1996). Probability and Information . Cambridge, UK: Cambridge University Press.

Barlow, H. B. (1962). A method of determining the overall quantum efficiency of visual discriminations. Journal of Physiology (London), 160, 155-168.

Barlow, H. B. (1977). Retinal and central factors in human vision limited by noise. In B. H. B., & F. P. (Ed.), Photoreception in Vertebrates Academic Press.

Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis . New York.: John Wiley & Sons.

Geisler, W. (1989). Sequential Ideal-Observer analysis of visual discriminations. Psychological Review, 96(2), 267-314.
Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. Proc Natl Acad Sci U S A.
Green, D. M., & Swets, J. A. (1974). Signal Detection Theory and Psychophysics . Huntington, New York: Robert E. Krieger Publishing Company.
Poor, H. V. (1994). An introduction to signal detection and estimation (2nd ed.). New York: Springer-Verlag.
Van Trees, H. L. (1968). Detection, Estimation and Modulation Theory . New York: John Wiley and Sons.
Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. 33, 113-120.