0.1 Likelihood Ratios, Decision Rules, and Anti-Neurons

(Notes for CNS 102, prepared by Jonathan Harel, Feb. 18. 2009)

Suppose we have two hypotheses h_1 and h_2 and only one observation of one event e_1. The best decision strategy would be to pick the hypothesis which is more probable given the observation. We will express this in formal symbols below.

Define:

$$\begin{aligned} p(h_i \text{ is true given observation } e) &\triangleq p(h_i \text{ is true } |e) = p(h_i|e) \\ p(\text{observing } e \text{ given that hypothesis } h_i \text{ is true}) &\triangleq p(e|h_i \text{ is true}) = p(e|h_i) \end{aligned}$$

Then, as we know from the Bayesian rules of probability:

$$\begin{aligned} p(h_1|e) &= \frac{p(e|h_1)p(h_1)}{p(e)} \\ p(h_2|e) &= \frac{p(e|h_2)p(h_2)}{p(e)} \end{aligned}$$

where $p(h_i)$ is the prior (before evidence) probability of h_i being true, and $p(e)$ is the probability with which you would observe e under either hypothesis. You can derive this arithmetic, namely $p(a,b) = p(a|b)p(b) = p(b|a)p(a)$, from a frequentist interpretation, in which what we are describing using these $p(\cdot)s$ are fractions of trials. We have some very large[1] number of trials, and we find that in a fraction $p(h_1)$ of the trials, h_1 is true, and in a fraction $p(h_2)$ of the trials, h_2 is true. $p(e)$ is the fraction of all trials in which e is observed. $p(e,h_1)$ is the fraction of all trials in which e is observed and h_1 is true, and $p(e|h_1)$ is the fraction of only those trials in which h_1 is true that e is observed, etc.. So, the denominator, which we actually will not use, can be expanded as:

$$p(e) \triangleq p(e|h_1)p(h_1) + p(e|h_2)p(h_2) = p(e,h_1) + p(e,h_2) = \sum_{i=1}^{2} p(e,h_i)$$

This kind of summation over some of the random variables which take on values in each trial (in this case h_i) is known as marginalization.

With perhaps enhanced intuition about these symbols, let's continue. h_1 is more probable than h_2, that is $p(h_1|e) > p(h_2|e)$, when the likelihood ratio $LR > 1$:

$$\text{likelihood ratio: } LR = \frac{p(h_1|e)}{p(h_2|e)} = \frac{p(e|h_1)p(h_1)}{p(e|h_2)p(h_2)}$$

[1] technically, as the number of trials increases, the fractions approach the $p(\cdot)$ values.

Our decision strategy is simply to select hypothesis h_1 if $LR > 1$, h_2 otherwise. The "1" can be substituted with any constant in general, if there are different costs for false alarms and misses[2]. This decision strategy can also be expressed in terms of the log likelihood ratio, LLR

$$LLR = \log(LR) = \log p(e|h_1) + \log p(h_1) - \log p(e|h_2) - \log p(h_2)$$

The strategy is: select h_1 as true if $LR > 1 \Leftrightarrow LLR > 0$ (or greater than any constant in general).

Now we will see what happens if instead of observing one event e_1, we also have another event e_2, which has a very peculiar **reciprocity property**, namely:

$$p(e_1 = X|h_1) = p(e_2 = X|h_2) \text{ and } p(e_1 = X|h_2) = p(e_2 = X|h_1). \tag{1}$$

That is, an observation at e_1 of X is exactly as probable under h_1 as it is under the *alternative* hypothesis h_2 when observed at e_2, and vice versa. For example, consider a neuron and an anti-neuron, tuned in exactly opposite ways. The observation of the first neuron's firing rate e_1 and the observation of the second, anti-neuron's, firing rate e_2 would obey this reciprocity property (1) if the following holds: the probability of the neuron firing at rate X under hypothesis 1 (e.g., "I am observing leftward motion"), is exactly equal to the probability of the anti-neuron firing at rate X under the *other* hypothesis (e.g., "I am observing rightward motion"), and vice versa.

In this case, the decision strategy derived from likelihood ratios will simplify. We start with the likelihood ratio:

$$LR = \frac{p(h_1|e_1 = X \text{ and } e_2 = Y)}{p(h_2|e_1 = X \text{ and } e_2 = Y)} = \frac{p(e_1 = X \text{ and } e_2 = Y|h_1)p(h_1)}{p(e_1 = X \text{ and } e_2 = Y|h_2)p(h_2)} \tag{2}$$

We stop here for an aside on an assumption which allows us to simplify the above expression to a more special case.

0.1.1 An aside on conditional independence

We assume that e_1 and e_2 are "conditionally independent" under each hypothesis. That is

$$p(e_1 \text{ and } e_2|h_i) = p(e_1|h_i)p(e_2|h_i) \text{ for } i \in \{1, 2\} \tag{3}$$

We just assume the probability of the pair of events factors in this way, which is the definition of conditional independence. Note that e_1 and e_2 are NOT in general independent[3] (i.e, $p(e_1 \text{ and } e_2) \neq p(e_1)p(e_2)$ in general), since if we

[2] If h_1 is "detection" and h_2 is "nothing there". In statistics, the claim that a decision rule of form $LR > k$ – for some constant not depending on the observations, k – is "most powerful" is known as the Neyman-Pearson lemma.

[3] An important note to make here is that, futhermore, independence does NOT guarantee conditional independence. In a frequentist interpretation, a hypothesis restricts us to some

know one event, we might be able to estimate which hypothesis is more likely, and that might tell us something about the other event, e.g. the other neuron's firing rate. They are only *conditonally independent* given each hypothesis. Explained in information theoretic terms, this means that if we *already know which hypothesis is true (the "given h" part)*, e.g., which direction the actual motion is, we don't gain any *additional information* about what might happen in one event after observing the other, e.g. the probability of one neuron's firing rate after observing the other's.

For instance, let's say that leftward motion means that a first neuron has a firing rate distributed Gaussian[4] around 100 spikes per second, and its anti-neuron has a firing rate distributed Gaussian around 25 spikes per second. Now let's say there is leftward motion. This what we already know. Now let's say that we observe 95 spikes per second on the first neuron. Do we modify our prediction for the anti-neuron? No, if the two events (firing rates) are conditionally independent, then the anti-neuron will still fire exactly according to the Guassian distribution around 25 spikes per second, not slightly modified in some way, e.g. according to the fact that the first neuron fired slightly below mean for leftward motion. The same would be true if the first neuron sporadically fired at 10 spikes per second.

Now let's continue our discussion of the likelihood ratio for two events obeying reciprocity property (1). Using conditional independence (3), we factor equation (2):

$$LR = \frac{p(e_1 = X|h_1)p(e_2 = Y|h_1)p(h_1)}{p(e_1 = X|h_2)p(e_2 = Y|h_2)p(h_2)}. \quad (4)$$

We will see what happens to this quantity (4) under different conditions.

0.1.2 Reciprocal events are equal in magnitude $(X = Y)$

If the two events are equal, that is if $X = Y$, then

$$LR = \frac{p(e_1 = X|h_1)p(e_2 = X|h_1)p(h_1)}{p(e_1 = X|h_2)p(e_2 = X|h_2)p(h_2)}$$

but we know from property (1) that $p(e_1 = X|h_1) = p(e_2 = X|h_2)$ and $p(e_2 = X|h_1) = p(e_1 = X|h_2)$. Thus

$$if\ e_1 = e_2 = X, \text{ then } LR = \frac{p(h_1)}{p(h_2)}$$

fraction of the trial set, and although averaged over the entire set, it may be the case that $p(x, y) = p(x)p(y)$, in some arbitrary subset, this factorization may break down, such that $p(x, y|z) \neq p(x|z)p(y|z)$. We might be able to carve out a subset of the trial set in which x is exactly equal to y, and say that's when $z = 1$. In this case, clearly x and y are not independent given z.

The bottom line is conditional independence and independence do not imply each other either way.

[4] or some nonnegative discrete analog of a Gaussian distribution, e.g., a binomial distribution.

This makes sense: in the neuron example, the neuron and its anti-neuron have produced the same firing rate to the stimulus – so what is the likelihood ratio? It is the same as our *prior* likelihood ratio, the one we arrive at, based on no evidence (events, observations) at all. Neuron and anti-neuron are responding equally, cancelling out any bias they might introduce about either hypothesis.

0.1.3 Reciprocal events have different magnitudes ($X \neq Y$)

We will simplify notation slightly. For the situation described above, there are only two probability distributions, which we will now label $f(\cdot)$ and $g(\cdot)$:

$$
\begin{aligned}
f(X) &= p(e_1 = X|h_1) = p(e_2 = X|h_2) \\
g(X) &= p(e_1 = X|h_2) = p(e_2 = X|h_1)
\end{aligned}
$$

Then, we will write the likelihood ratio for two conditionally independent events in terms of these:

$$
\begin{aligned}
LR &= \frac{p(e_1 = X|h_1)p(e_2 = Y|h_1)p(h_1)}{p(e_1 = X|h_2)p(e_2 = Y|h_2)p(h_2)} \\
&= \frac{f(X)g(Y)}{g(X)f(Y)} \cdot \frac{p(h_1)}{p(h_2)}
\end{aligned}
$$

Assume for simplicity that each hypothesis has the same prior probability so that

$$\frac{p(h_1)}{p(h_2)} = 1$$

And compute the log likelihood ratio:

$$LLR = \log LR = \log f(X) - \log g(X) - (\log f(Y) - \log g(Y))$$

Now we define a new function

$$k(X) \triangleq \log f(X) - \log g(X)$$

then:

$$LLR = k(X) - k(Y).$$

So for equal prior probabilities, and equal treatment of false alarms and misses (comparison constant=0), the optimal decision rule is "h_1 is more likely given the two observations if $LLR > 0$, which is true if and only if $k(X) - k(Y) > 0$".

This takes an even simpler form if f and g are Gaussians of the same standard deviation but different means a and b:

$$
\begin{aligned}
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-a)^2/2\sigma^2) \\
g(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-b)^2/2\sigma^2)
\end{aligned}
$$

$$\begin{aligned}
k(x) &= \log f(x) - \log g(x) \\
&= -(x-a)^2/2\sigma^2 + (x-b)^2/2\sigma^2 \\
&= \frac{1}{2\sigma^2}\left(-(x^2 - 2ax + a^2) + (x^2 - 2bx + b^2)\right) \\
&= \frac{1}{2\sigma^2}(2ax - 2bx - a^2 + b^2) \\
&= \frac{x}{\sigma^2}(a-b) + \frac{b^2 - a^2}{2\sigma^2}
\end{aligned}$$

Which means, in this case:

$$\begin{aligned}
LLR &= k(x) - k(y) = \frac{x}{\sigma^2}(a-b) + \frac{b^2-a^2}{2\sigma^2} - \frac{y}{\sigma^2}(a-b) - \frac{b^2-a^2}{2\sigma^2} \\
&= (x-y)\frac{(a-b)}{\sigma^2}
\end{aligned}$$

Thus, if the sign of $a - b$ is known, then the optimal decision (guess h_1 is true if $LLR > 0$) can be made using only the sign of the difference between reciprocal observations x and y.

So, perhaps, if there are reciprocally tuned pairs of neurons in the brain, one can imagine a neural mechanism for an organism optimally deciding between two hypotheses. The decision outcome would depend on the spiking rate of a third neuron, which computes which neuron – primary or anti-neuron – is firing more rapidly.