Perceptual Distance and Visual Search

Data Science - Visual Neuroscience Lecture 3
A quick recapitulation

- We are trying to quantify perceptual distance between objects.

- Two different ways and a comparison.
  - Via behavioural experiments for detecting an oddball among distracters.
  - By capturing neuron responses.

- Towards this, we looked at a simplified model with true state of nature being one image or the other, and a single neuron observation.

- Hypothesis testing, model for observations as points of a Poisson point process, optimality of likelihood ratio test, log-likelihoods viewed as (random) information, the additivity property of log-likelihoods, its expectation is relative entropy under one hypothesis (positive) and negative relative entropy under the other (negative).

- Relative entropy as a measure of dissimilarity between two probability distributions. Data processing inequality.
Suppose policy $\pi$ says “no matter what, stop at $T$”.

$x = (a_1, x_1, a_2, x_2, \ldots, a_{T-1}, x_{T-1}, a_T = \text{stop})$.

By additivity of log-likelihoods

$$D(P_0^\pi \| P_1^\pi) = E_0 \left[ \sum_{t=1}^{T} \log \frac{p_0(X_t)}{p_1(X_t)} \right] = TD(p_0 \| p_1),$$

where $D(p_0 \| p_1)$ is relative entropy of 1 sample.

But we are interested in a stopping rule that depends on the observations.

A result from probability theory: Optional stopping theorem (without proof)

$$D(P_0^\pi \| P_1^\pi) = E_0 \left[ \sum_{t=1}^{\tau} \log \frac{p_0(X_t)}{p_1(X_t)} \right] = E_0[\tau]D(p_0 \| p_1).$$
\( D(Q_0^{\pi} \parallel Q_1^{\pi}) \), and a summing up

- **Interpretation of \( Q_0^{\pi} \):** Under hypothesis \( H_0 \), when you stop, probabilities of various decisions

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Distribution</th>
<th>Decision 0</th>
<th>Decision 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 )</td>
<td>( Q_0^{\pi} )</td>
<td>( \geq 1 - \epsilon )</td>
<td>( \leq \epsilon )</td>
</tr>
<tr>
<td>( H_1 )</td>
<td>( Q_1^{\pi} )</td>
<td>( \leq \epsilon )</td>
<td>( \geq 1 - \epsilon )</td>
</tr>
</tbody>
</table>

- Approximately \( D(\{1 - \epsilon, \epsilon\} \parallel \{\epsilon, 1 - \epsilon\}) \)

\[
(1 - \epsilon) \log \frac{1 - \epsilon}{\epsilon} + \epsilon \log \frac{\epsilon}{1 - \epsilon} \sim \log \frac{1}{\epsilon}.
\]

- **Thus:** \( E_0[\tau]D(p_0 \parallel p_1) \gtrsim \log \frac{1}{\epsilon} \), or

\[
E_0[\tau] \gtrsim \frac{\log \left( \frac{1}{\epsilon} \right)}{D(p_0 \parallel p_1)}.
\]
Is there a policy that will achieve this?

- Yes, asymptotically ... (Wald, late 1940s.)

- Accumulate $\log \frac{p_0(x_t)}{p_1(x_t)}$ across time. Wait until it exceeds a high enough threshold.

- Trade-off between confidence and delay.

- Lower bound suggests that we should stop at $\log(1/\varepsilon)$. This is the same as likelihood ratio $\frac{P_{\pi}^0(\cdots)}{P_{\pi}^1(\cdots)} \geq \frac{1}{\varepsilon}$. This is what makes it an $\varepsilon$-admissible policy.

- Policy:
  - Start with $S_0 = 0$.
  - At time $t$, compute $S_t = S_{t-1} + \log \frac{p_0(x_t)}{p_1(x_t)}$.
  - If $S_t > \log(1/\varepsilon)$, stop and decide $H_0$.
    If $S_t < -\log(1/\varepsilon)$, stop and decide $H_1$.
    Otherwise, continue.
A candidate for perceptual distance

- Search times are proportional to $\frac{1}{D(p_0||p_1)}$.

- If subjects wait to gather the same degree of confidence, then

  \[
  \frac{D(p_0||p_1)}{N} = \text{perceptual distance between image 0 and image 1.}
  \]

  \[N = \text{number of neurons under consideration.}\]

- A simple calculation yields:

  \[
  D(p_0||p_1) = \sum_n \left[ \lambda_0(n) \log \frac{\lambda_0(n)}{\lambda_1(n)} - \lambda_0(n) + \lambda_1(n) \right].
  \]

- Oddball image is $i$ and distractor is $j$, then $D(p_i||p_j)/N =: D_{ij}$. 

Search with control

- We actually have controls as well. Which place to look at.

- A more detailed model with controls provides us with a refinement. We will not go into the details here. But you have a homework question.

- But instead, we will stick to $D_{ij}$ for the data analysis.
Other natural distance candidates?

- Another proposal: $L_{ij} = N^{-1}||\lambda_i - \lambda_j||_1 = \frac{1}{N} \sum_n |\lambda_0(n) - \lambda_1(n)|$.

- Symmetric.

- This has a drawback, because we know that $Q$ in a sea of $O$'s is easier to identify that $O$ in a sea of $Q$'s.
Estimating relative entropy

- We don’t really know the true firing rates. We estimate them based on firing rate measurements, which are noisy.

- If we plug in the estimated rates into the formula for relative entropy, we will suffer a bias.

- The expected value of

\[ \hat{\lambda}_0 \log(\hat{\lambda}_0/\hat{\lambda}_1) - \hat{\lambda}_0 + \hat{\lambda}_1 \]

can be different from the true value for different \((\lambda_0, \lambda_1)\) pairs.

- You should try:

\[ \hat{D}_{01} = \begin{cases} 
\left[ \hat{\lambda}_0 \log \frac{\hat{\lambda}_0 - 1/(2m\Delta)}{\hat{\lambda}_1 + 1/(2m\Delta)} - \hat{\lambda}_0 + \hat{\lambda}_1 \right]_+ & \text{if } \hat{\lambda}_0 > 1/(2m\Delta), \\
\hat{\lambda}_1, & \text{otherwise}.
\]

\(m = 24, \Delta = 250\) ms from the Sripati and Olson experiments.
Assignment: Correlation analysis

- Divide data into groups. Each group is for an ordered image pair.

- Compute $s_{ij}$, $\hat{D}_{ij}$, $L_{ij}$. $s_{ij}$ plays the role of $\tau$.
  Remember to subtract the baseline reaction time of 328 ms to get time for decision alone.
  Remember to treat the compound searches correctly.

- Given $(s_{ij}^{-1}, \hat{D}_{ij})$, find the best straight line passing through the origin.
  Given $(s_{ij}^{-1}, L_{ij})$, find the best straight line passing through the origin.

- Which gives a better fit?
With the more refined perceptual distance

\[ \text{Neuronal Metric } \tilde{D}_i \]

\[ \text{Behavioural index } (s^{-1}) \]

**Behavioural index and proposed neuronal metric**

\[ r = 0.94378 \quad p = 4.707e^{-12} \]
Assignment: A measure of spread

- What we anticipate is that
  \[ u_{ij} := s_{ij} \times \hat{D}_{ij} \sim \text{constant, across } i, j. \]

- Similarly,
  \[ v_{ij} := s_{ij} \times L_{ij} \sim \text{constant, across } i, j. \]

- Which fits the observations better?

- A measure of spread is AM/GM of the \( u_{ij} \)'s and the \( v_{ij} \)'s.

- Higher this ratio, greater the spread.
Assignment: Guessing the distribution of the search times

- We did not cover this in class, but you will do it in your assignment.
- Pick (randomly) half of the groups and get a scatter plot of the (mean, stddev).
- You will see that stddev is roughly proportional to the mean.
- Fit a Gamma distribution which has this property.
- Density is \( g(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \), \( x \geq 0 \).
  \( \alpha \) is the shape, \( \beta \) is the rate.
- Mean = \( \frac{\alpha}{\beta} \), stddev = \( \sqrt{\frac{\alpha}{\beta}} \), so that stddev/mean = \( \frac{1}{\sqrt{\alpha}} \).
  Fit a straight line to the scatter plot above and provide a guess for the shape \( \alpha \).
The Kolmogorov-Smirnov statistic

- On each of the groups that did not contribute to your shape parameter, randomly select one half of the data points and estimate the rate parameter.

- Plot the cdf with the estimated shape and rate and call it $F(x)$.

- Plot the cdf of the remaining data in the group. Let the samples be $s(1), s(2), \ldots, s(K)$.

$$\hat{F}(x) = \frac{1}{K} \sum_{k=1}^{K} 1\{s(k) \leq x\}.$$

This is the empirical cdf.

- How close are the two? What is the max distance between the first and the second cdfs?

$$KS = \max_x |F(x) - \hat{F}(x)|$$
Assignment: Hint on the general case

Consider two hypotheses $h$ and $h'$.

Let $A_t$ be the action at time slot $t$. Let $N_a(t)$ be the number of times $a$ is chosen in slots upto $t$.

$$D(P^\pi_h || P^\pi_{h'}) = E^\pi_h \sum_{t=1}^{\tau} \log \frac{p^{A_t}_h(X_t)}{p^{A_t}_{h'}(X_t)} \quad \text{(conditional independence)}$$

$$= E^\pi_h \sum_{a=1}^{K} \sum_{l=1}^{N_a(\tau)} \log \frac{p^a_h(X_l)}{p^a_{h'}(X_l)}$$

$$= \sum_{a=1}^{K} E^\pi_h [N_a(\tau)] D(p^a_h || p^a_{h'}) \quad \text{(Optional stopping)}$$

$$\leq E^\pi_h [\tau] \max_{\lambda} \sum_{a=1}^{K} \lambda_a D(p^a_h || p^a_{h'}).$$

How should an adversary choose $h'$ to minimise the information content in each slot? How should the searcher choose $\lambda$ to maximise his information content?
What did we learn in this module?

- Hypothesis testing
- Hypothesis testing with a stopping criterion
- Relative entropy
- Data processing inequality
- Some asymptotic analysis
- Fitting a distribution, Kolmogorov-Smirnov statistic
- A measure of spread AM/GM.