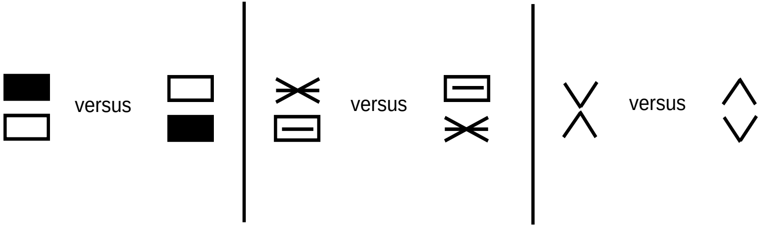
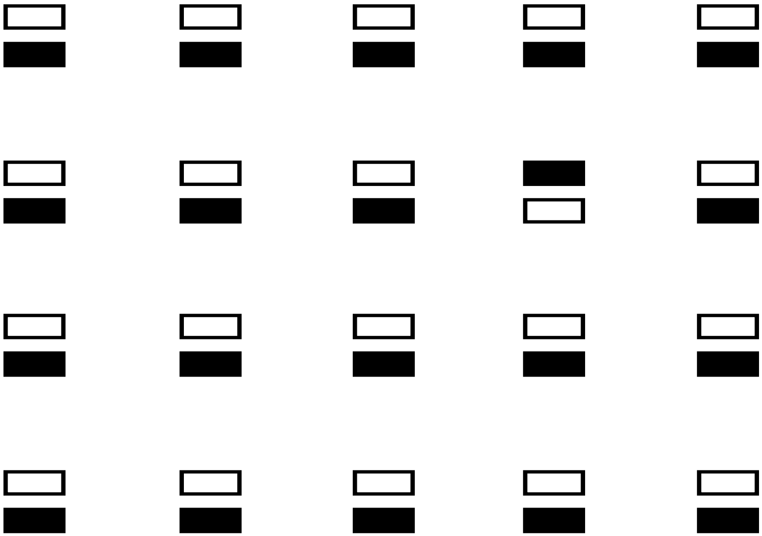
Perceptual Distance and Visual Search

Data Science - Visual Neuroscience Lecture 2

Measuring perceptual distance



Find the odd image

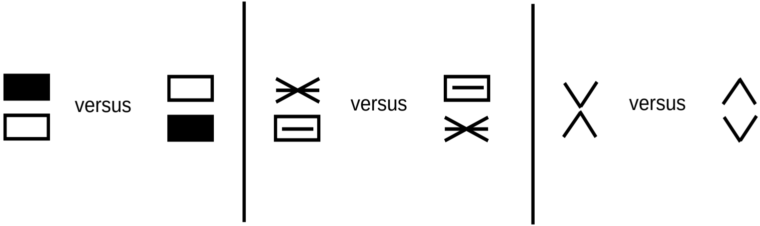


A measure of perceptual distance

Hypothesis

Visual search performance depends on the perceptual distance between the two images. Closer the two images in perceptual distance, the longer it takes to identify the oddball image. More specifically:

$$\text{Proposed Perceptual Distance} \propto \frac{1}{(\text{Search Time})^k}?$$



From the IT of the macaques

- Inferotemporal cortex - gross object features emerge here
- Firing rates of $N = 174$ neurons in response to these six images
- Data collected in a similar manner for a total of 24 images
- For each image i, the neuronal response is summarized by the firing rate vector $(\lambda^i(n), 1 \leq n \leq N)$.

$$\text{Image } i \mapsto \lambda^i = \begin{pmatrix} \lambda^i(1) \\ \lambda^i(2) \\ \vdots \\ \lambda^i(N) \end{pmatrix}$$

The main question

- For the pair (i, j), perceptual distance ought to be a function of how "different" λ^i and λ^j are.
- What function?
- How does it relate to reaction time?

A model grounded in a theory

- What would the prefrontal cortex do if it got observations from the human analogue of the inferotemporal cortex and could control the eye?

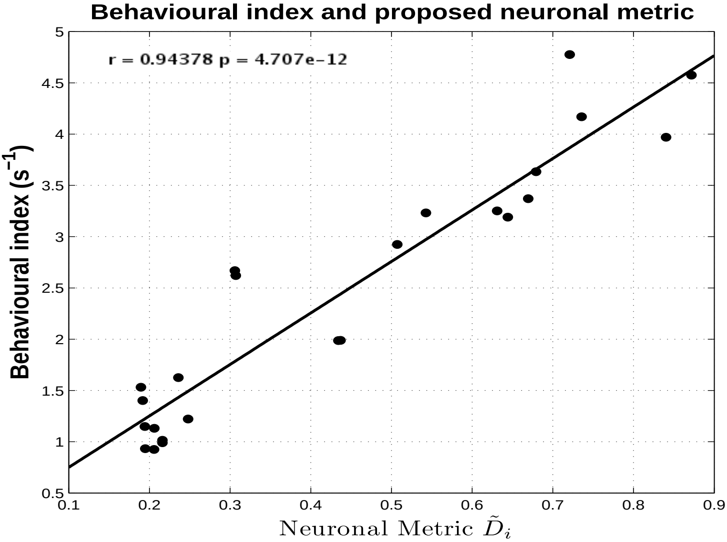
A model for search - sequential hypothesis testing

- Hypothesis $h = (\ell, i, j)$: The oddball location is ℓ and its type i among distracters j. Ground truth.
- Divide time into slots.
- Control: Given observations and decisions in all previous slots (history),
  - decide to stop and declare the oddball, or
  - decide to continue, and direct the eye to focus on location b, one of the six locations.
- Observation: If the object in location b is k, then N Poisson point processes with rates $(\lambda^k(n), 1 \leq n \leq N)$.
- Policy π: For each time slot, given history, a prescription for action.
  To stop or not to stop?
  If continue, where to look?
  If stop, what to decide?

Performance

- For each ground truth h, your policy shall make an error with probability at most ε.
- What is the expected time to stop for a fixed positive ε?
- The average search delay is the average over all hypotheses h with i as oddball and j as distracter.
- What function of λ^i and λ^j?
  Difficult to evaluate. We will do some asymptotics as $\varepsilon \to 0$ to get the following.

We will process data to get this correlation plot

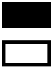


What we will learn in this module

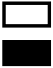
- Hypothesis testing
- Hypothesis testing with a stopping criterion
- Data processing inequality, and relative entropy
- A brief view into asymptotic analysis
- Testing for a distribution - Kolmogorov-Smirnoff test
- ANOVA and variants

A much simplified hypothesis testing problem

- Suppose only two states of nature.
- Either the picture is



  or the picture is



- Call the first H_0 and the second H_1.
- You get to look at it for one second. You have to decide if H_0 or H_1.
- Limitation. You have only one neuron. $N = 1$.
- If the true state of nature were H_0, the neuron fires at rate λ_0.
  If the true state is H_1, the neuron fires at rate λ_1.
- Observe X spikes. If you see $X = 5$ spikes, which image?

Poisson point process of rate λ

- This is an often used simplified model for spike trains.

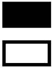
- Properties:
  - If A and B are two disjoint sets, then the number of points X_A and X_B in A and B are independent random variables.
  - If A has size (length) $m(A)$, then the expected number of points $E[X_A] = m(A)$.
- Poisson point process of rate λ: Expected number of points in an interval of length 1 is λ.
- This suffices to describe the process completely. We can deduce that the number of points X in $[0, 1]$ has the Poisson distribution:

$$\Pr\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \geq 0.$$

- Proof: Binomial converges to the Poisson distribution when scaled appropriately.

The distributions under the two hypothesis

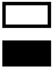
- When the picture is



X has distribution Poisson(λ_0),

$$\Pr\{X = k|H_0\} = p_0(k) = (\lambda_0)^k e^{-\lambda_0}/k!.$$

- Similarly, when the picture is



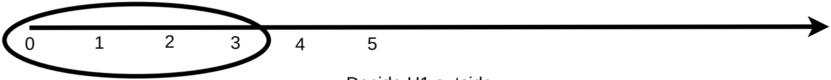
X has distribution Poisson(λ_1),

$$\Pr\{X = k|H_1\} = p_1(k) = (\lambda_1)^k e^{-\lambda_1}/k!.$$

- What if you have N neurons?

Decision rule, performance criterion

- Decision rule: $\delta : \{0, 1, 2, \ldots\} \to \{H_0, H_1\}$



- Partition observation space into Γ_0 (decide H_0) and Γ_1 (decide H_1).
- Performance criterion: Probability of error
- Assume each hypothesis is equally likely. Then

$$\Pr\{\text{Error}\} = \frac{1}{2}\Pr\{\delta(X) = H_1 | H_0\} + \frac{1}{2}\Pr\{\delta(X) = H_0 | H_1\}.$$

- Choose the decision rule that minimises probability of error.

Likelihood ratio test

- Fact: Assume H_0 and H_1 are equally likely. The optimal decision rule that minimises the probability of error is the following.

$$\delta(x) = \begin{cases} H_1 & \text{if } p_1(x) > p_0(x) \\ Either & \text{if } p_1(x) = p_0(x) \\ H_0 & \text{otherwise.} \end{cases}$$

- Proof: Think of $\delta(x) \in [0, 1]$ as a probability assignment for a randomised decision:

$$\begin{aligned} \Pr\{\text{error}\} &= \frac{1}{2} \sum_{x \geq 0} p_0(x)\delta(x)\ dx + \frac{1}{2} \sum_{x \geq 0} p_1(x)[1 - \delta(x)]\ dx \\ &= \frac{1}{2} + \frac{1}{2} \sum_{x \geq 0} [p_0(x) - p_1(x)]\ \delta(x)\ dx. \end{aligned}$$

  For each x, choose $\delta(x)$ to make the integrand as small as possible.

- Same as $\frac{p_1(x)}{p_0(x)}$ being compared with 1,
  or equivalently, $\log \frac{p_1(x)}{p_0(x)}$ being compared with 0.

Relative entropy

- Working with the log. Suppose we have observations in two slots, say x_1, x_2.

  Log likelihood ratio $= \log \frac{p_1(x_1)}{p_0(x_1)} + \log \frac{p_1(x_2)}{p_0(x_2)}$ is additive in the observations.

- Expectation of the log likelihood:

$$E_0\left[\log \frac{p_1(X)}{p_0(X)}\right] \text{ and } E_1\left[\log \frac{p_1(X)}{p_0(X)}\right]$$

- Relative entropy of p with respect to q, denoted $D(p||q)$, is defined as

$$D(p||q) = E_p\left[\log \frac{p(X)}{q(X)}\right] = \sum_{x \geq 0} p(x) \log \frac{p(x)}{q(x)}.$$

- Fact: $D(p||q) \geq 0$ with equality if and only if $p = q$.
  A measure of how far apart p and q are from each other.
  Asymmetric!

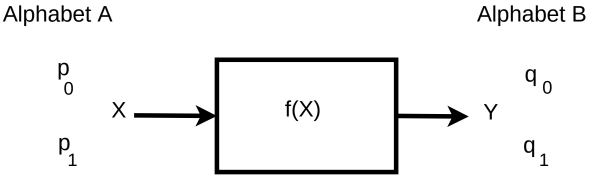
Proof: $D(p||q) \geq 0$ with equality if and only if $p = q$

- This is the same as showing $-D(p||q) = \sum_{x \geq 0} p(x) \log \frac{q(x)}{p(x)} \leq 0$.
- A useful inequality: $\log u \leq u - 1$ for all $u \geq 0$ with equality if and only if $u = 1$. Natural logarithm.
- Substitute, and be a little more careful:

$$
\begin{aligned}
-D(p||q) &= \sum_{x:p(x)>0} p(x) \log \frac{q(x)}{p(x)} \\
&\leq \sum_{x:p(x)>0} p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \\
&= \sum_{x:p(x)>0} (q(x) - p(x)) \\
&= Q(supp(P)) - 1 \\
&\leq 0.
\end{aligned}
$$

- Condition for equality is easy now.

A data processing inequality



- Observe X, test for H_0 versus H_1 on the left.
  Process X. Now keep only Y. Test for H_0 versus H_1 on the right.
- $A = \mathbb{Z}_+$. $y = f(x) = 1\{x \geq 20\}$. What is the alphabet B? q_0? q_1?
- Fact: The data processing inequality $D(p_0||p_1) \geq D(q_0||q_1)$ holds.

Proof of data processing inequality

- $LHS = \sum_{x \geq 0} p_0(x) \log \frac{p_0(x)}{p_1(x)}$, $\quad RHS = \sum_{y \in B} q_0(y) \log \frac{q_0(y)}{q_1(y)}$.
- Fix y. Take $f^{-1}(y) = \{x \geq 0 | f(x) = y\}$.
  $q_0(y) = \sum_{x \in f^{-1}(y)} p_0(x)$.
- Focus on this y and the corresponding terms on the LHS.
- Claim: $\sum_i a_i \log \frac{a_i}{b_i} \geq a_{sum} \log \frac{a_{sum}}{b_{sum}} = \sum_i a_i \log \frac{a_{sum}}{b_{sum}}$.
- This is the same as

$$
\begin{aligned}
\sum_i a_i \left[\log \frac{a_i}{b_i} - \log \frac{a_{sum}}{b_{sum}} \right] &\geq 0 \\
\sum_i a_i \log \frac{a_i / a_{sum}}{b_i / b_{sum}} &\geq 0 \\
\sum_i (a_i / a_{sum}) \log \frac{a_i / a_{sum}}{b_i / b_{sum}} &\geq 0.
\end{aligned}
$$

  This holds because the left side is a relative entropy.

Hypothesis testing with a stopping criterion: policy

- In the one sample likelihood ratio test, probability of error is whatever you get.
- What if we want a target probability of error?
- Two approaches:
  Up front decide on how many slots to view. Fixed sample size.
  Continue to view until you meet target error probability criterion: policy
- Policy π: at the beginning of each slot, given past observations and actions,
  - decide to stop and declare H_0 or H_1
  - decide to continue.
  - Can think of $\pi = (\pi_1, \pi_2, \ldots)$, where
  - $(a_1, x_1, a_2, x_2, \ldots, a_{t-1}, x_{t-1}) \mapsto \pi_t(\cdots) = a_t \in \{stop,\ continue\}$, and
  - when stop, $\delta(\cdots) \in \{H_0, H_1\}$.
- Notation: $P_0^\pi(Event) = \Pr\{Event\ |H_0,\ \text{policy is } \pi\}$.

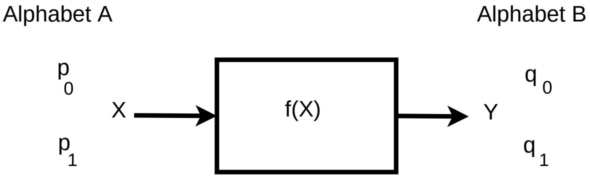
Hypothesis testing with a stopping criterion: performance criteria

- Performance criterion 1: We say that a policy π is ε-admissible if both
$$P_0^\pi\{\delta(\cdots) \neq H_0\} \leq \varepsilon \text{ and } P_1^\pi\{\delta(\cdots) \neq H_1\} \leq \varepsilon.$$

- Performance criterion 2: Let τ be the stopping time
$$\tau := \min\{t \geq 1 | \pi_t(\cdots) = \textit{ stop}\}.$$
Expected stopping times: $E_0^\pi[\tau], E_1^\pi[\tau], (E_0^\pi[\tau] + E_1^\pi[\tau])/2$.

- Minimise expected time to stop among all ε-admissible policies.

Data processing inequality again, and a homework



- Consider P_0^π and P_1^π. Similarly for Q.
  Let $x = (a_1, x_1, a_2, x_2, \ldots, a_{\tau-1}, x_{\tau-1}, a_\tau = stop, \delta)$.
  Let $y = \delta$.
- The data processing inequality is $D(P_0^\pi || P_1^\pi) \geq D(Q_0^\pi || Q_1^\pi)$.
  If π is ε-admissible, what happens to the right-hand side as $\varepsilon \to 0$.