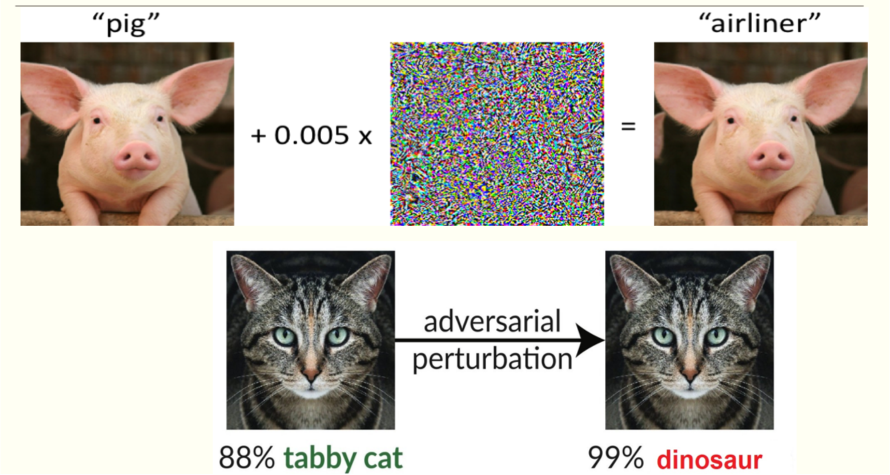
Perceptual Distance and Visual Search

Data Analytics - Visual Neuroscience Lecture 1

Robust visual perception



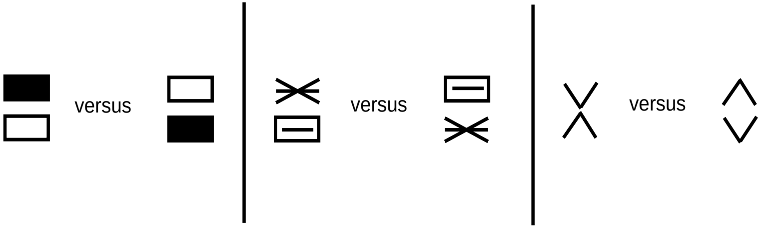
Kent Face Matching Test

Top pair - match, bottom pair - mismatch

Perceptual distance

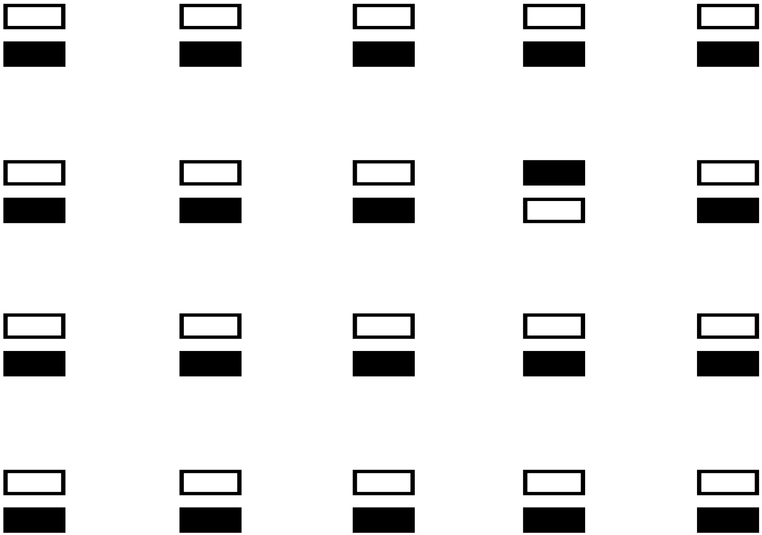
- How are objects represented in our brains to enable us to tell one from another?
- "Distance" between objects, as represented in our brains
- "Pixel distance" very different from "perceptual distance"
- In this module: Study experimental data that attempts to quantify perceptual distance

Measuring perceptual distance

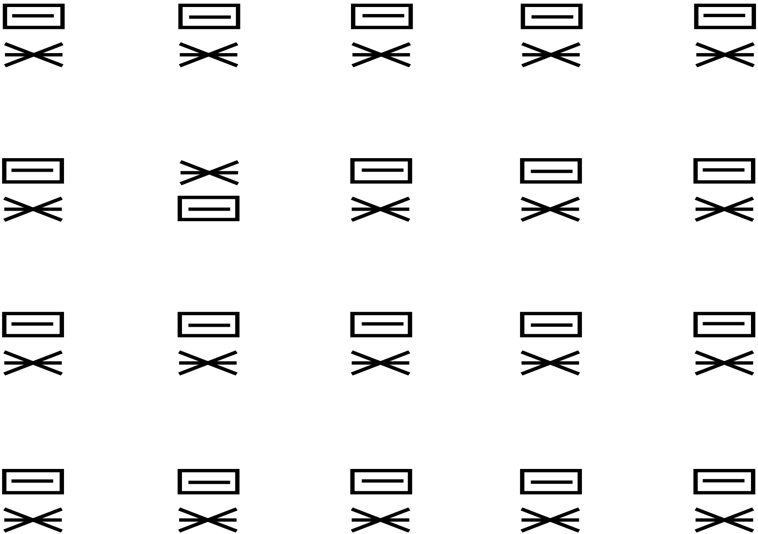


Ideas?

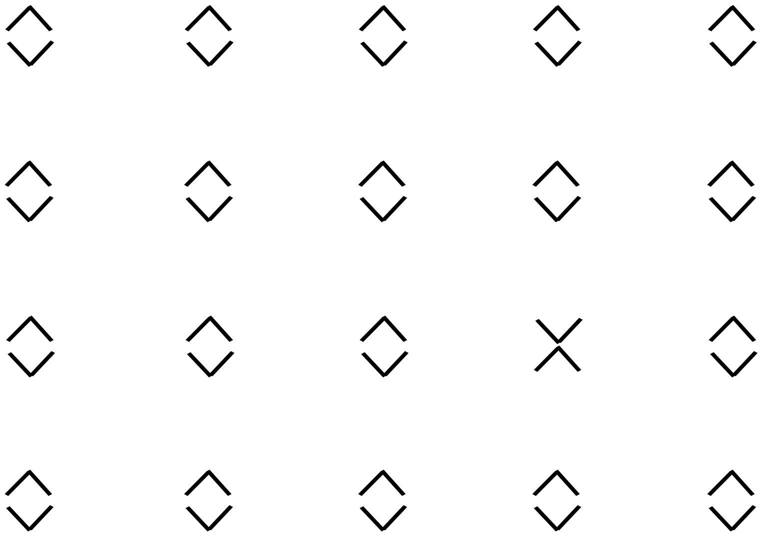
Find the odd image - 1



Find the odd image - 2



Find the odd image - 3

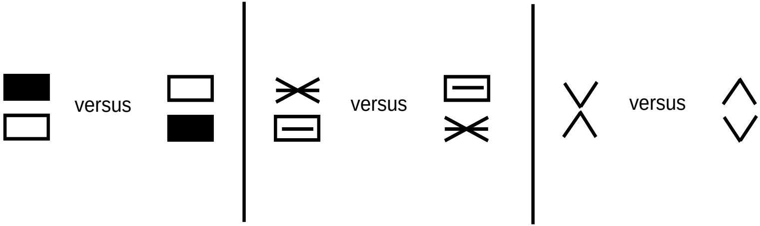


A measure of perceptual distance

Hypothesis

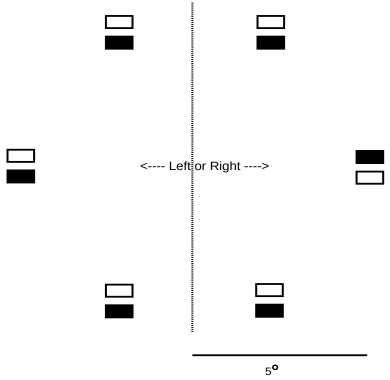
Visual search performance depends on the perceptual distance between the two images. Closer the two images in perceptual distance, the longer it takes to identify the oddball image. More specifically:

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



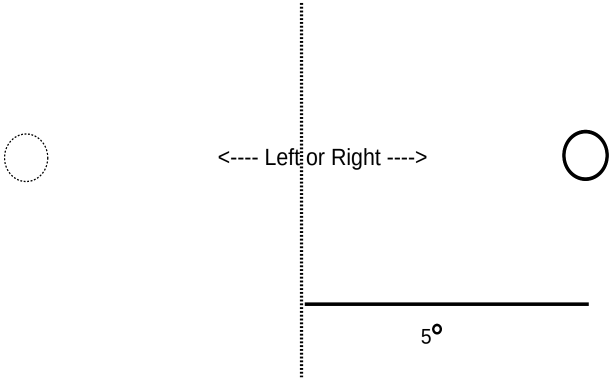
A reaction time study on humans (Arun and Olson 2010)

- Study conducted on six subjects
- Identify the location of the oddball and hit a key to tell left or right



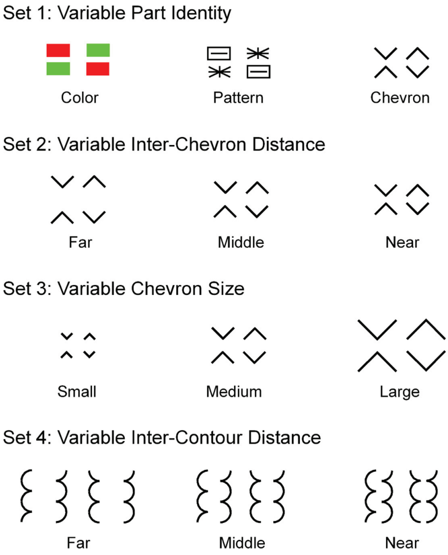
- Image displayed until reaction (which if correct, valid trial), or until 5 seconds (aborted)
  $\mathrm{RT}(i, j)$ = average reaction time
  Data averaged over both *oddball i among distracters j*

Baseline reaction time



- RT_b = baseline reaction time
- $s(i,j) = \mathrm{RT}(i,j) - \mathrm{RT}_b$
- Perceptual distance between i and j is $\propto 1/s(i,j)$
- $\mathrm{RT}_b = 328ms$.

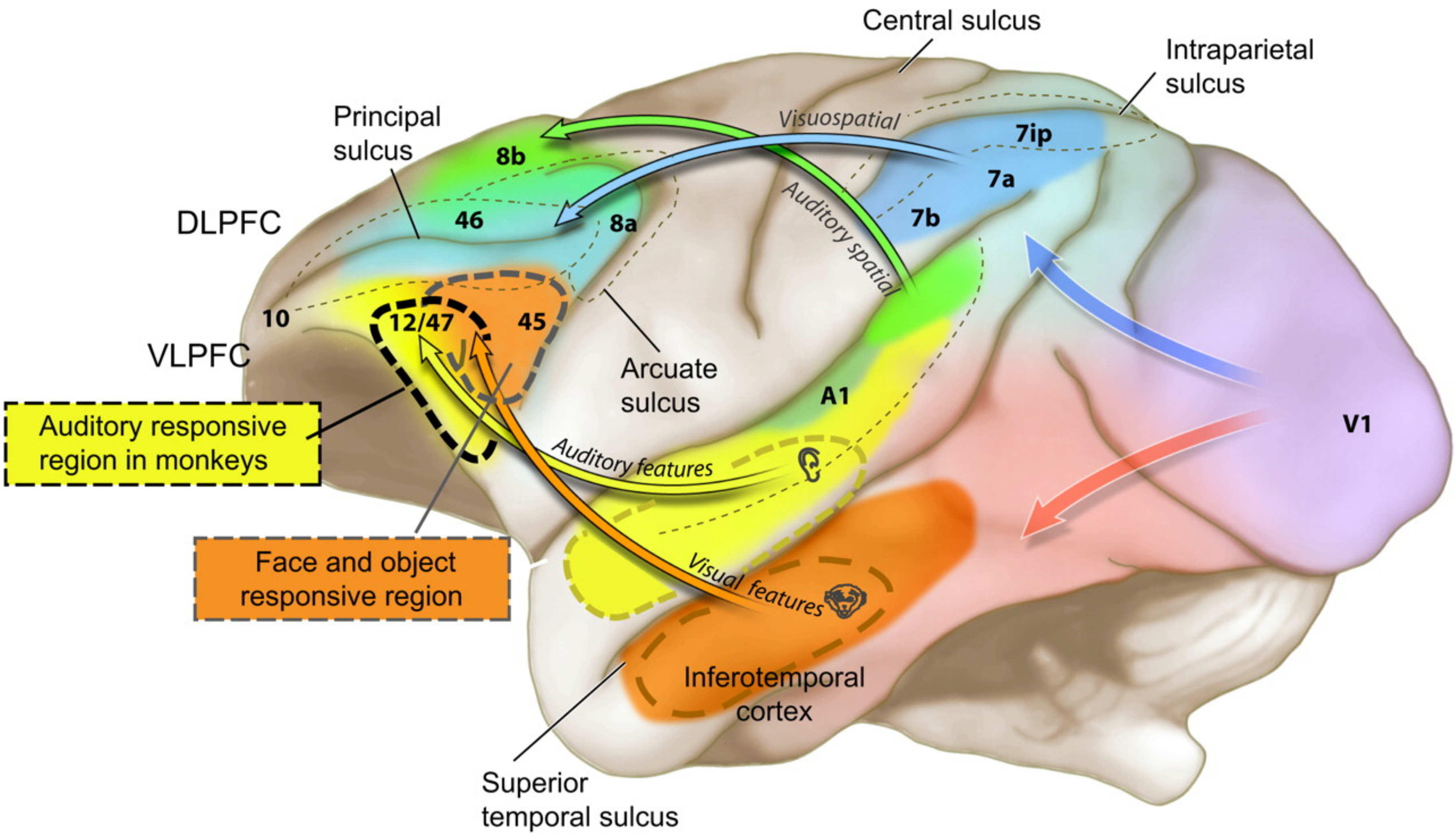
Image pairs on which search time data was collected (Sripati and Olson 2010)



A direct view into the brain of rhesus macaques

- Try to nail the responses in the brain, and see how different they are.

A schematic of the macaque brain is shown depicting the flow of auditory and visual information through the brain to its ultimate destination in PFC. Regions of the prefrontal cortex are color coded to match those of areas, that project to it.



Romanski L M Cereb. Cortex 2007;17:i61-i69

Cerebral CORTEX

Where to measure?

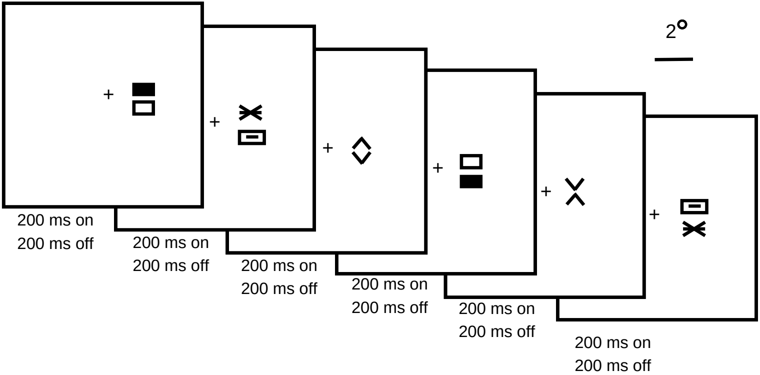
- The case for measuring in IT (Sripati and Olson 2010)
- Neurons in IT, unlike those in low-order visual areas, have receptive fields large enough to capture an entire image.
- Sensitive to global arrangement of elements within the image.
- Studies indicate that population activity in IT discriminates some images better than others. Studies also indicate that if a pair is well-discriminated by population activity in IT, then humans tend to characterise them as dissimilar.
- Perhaps then population activity in IT should predict human search efficiency.

Experimental procedure on rhesus macaques and recording (done by Arun and Olson)

- Cleared by CMU institutional animal care and use committee
- Two macaques were surgically fitted with:
  - a cranial implant for neuronal activity recording;
  - a scleral search coil for recording eye movements.
- Data was collected over several days. Before each day's experiment, an electrode was inserted so that the tip was 1 cm above the inferotemporal cortex.
- The electrodes were pushed, reproducably, along tracks forming a square grid with 1 mm spacing.
- Neuronal activity was recorded. Individual neurons' action potentials then isolated using a commercially available tool (Plexon).

A direct view into the brain of rhesus macaques

- Two macaques were trained to fixate on the $+$ while a series of stimuli appeared one after another.



- Images were randomly interleaved. Neuronal activity recorded (inferotemporal cortex) over several 2 second rounds.

The neuronal data

- 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?

Aspects of search

- Find in the shortest possible time. Cost = delay.
- Local focus. You could choose where you wanted to look next.
- Two types of pictures. But you didn't "know" either. Learnt which is which on the fly.
- But you learnt just enough to tell a picture in location 1 was same as or different from the picture in location 2.
- When you changed focus, you often chose a location nearer to the current location.
- You waited until you were sure about the oddball location.

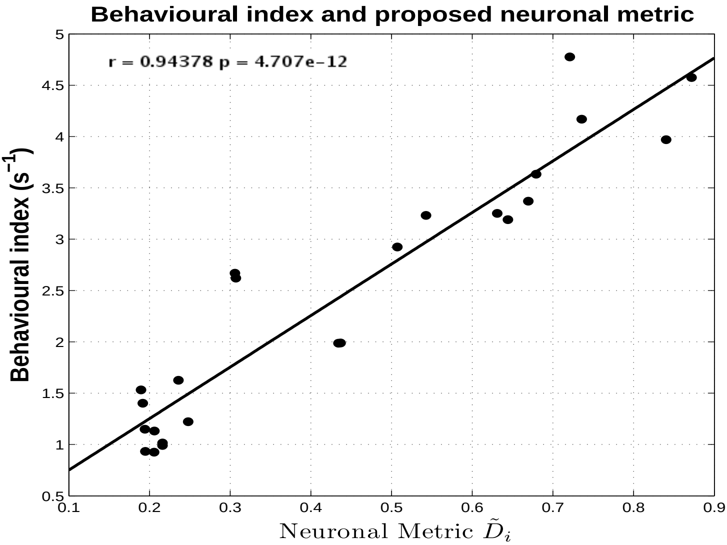
A model for search - sequential hypothesis testing

- 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



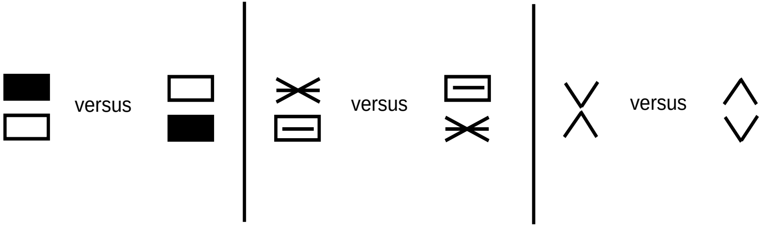
What we will learn in this module

- 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

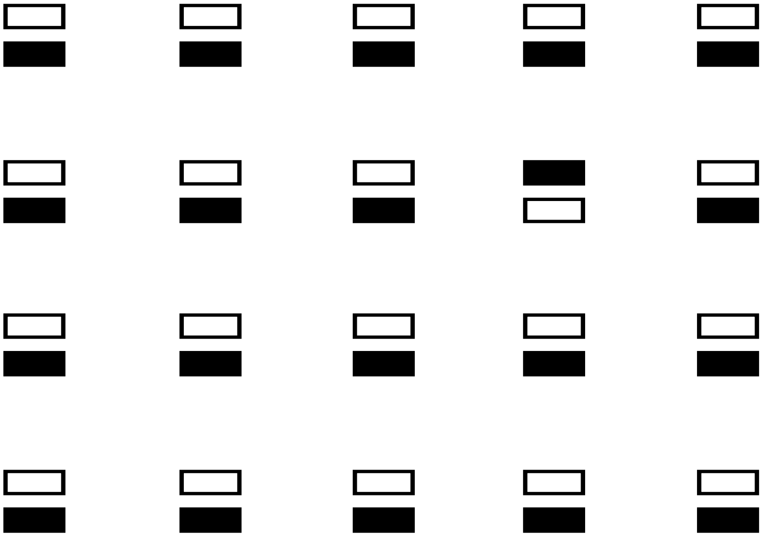
Perceptual Distance and Visual Search

Data Analytics - Visual Neuroscience Lecture 2

Measuring perceptual distance



Find the odd image

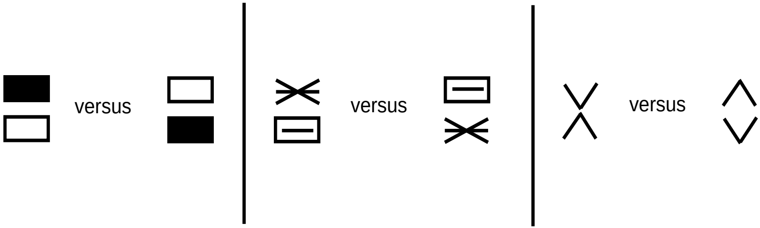


A measure of perceptual distance

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

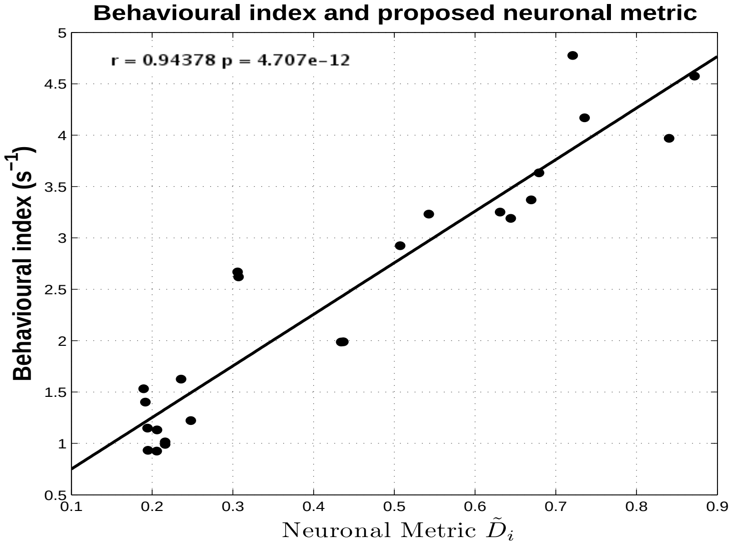
- For each ground truth h, your policy shall make an error with probability at most ε.

- 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

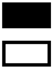


What we will learn in this module

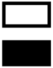
- Hypothesis testing
- Hypothesis testing with a stopping criterion
- Data processing inequality, and relative entropy
- A brief view into asymptotic analysis
- Testing for a distribution - Kolmogorov-Smirnoff test
- ANOVA and variants

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.

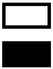
The distributions under the two hypothesis

- 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



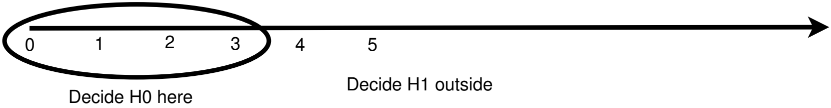
  X has distribution Poisson(λ_1),

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

- 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!

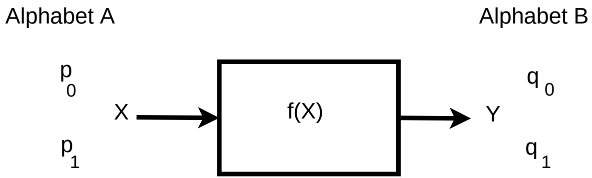
Proof: $D(p||q) \geq 0$ with equality if and only if $p = q$

- 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: $P_0^{\pi}(Event) = \Pr\{Event\ |H_0,\ \text{policy is } \pi\}$.

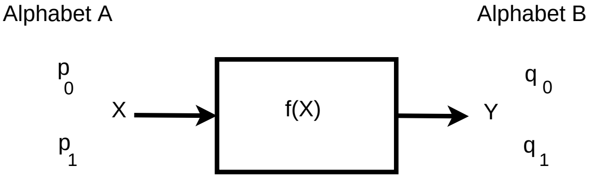
Hypothesis testing with a stopping criterion: performance criteria

- 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 $x = (a_1, x_1, a_2, x_2, \dots, a_{\tau-1}, x_{\tau-1}, a_\tau = \mathit{stop}, \delta)$.
  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 $\varepsilon \to 0$.

Perceptual Distance and Visual Search

Data Analytics - 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.

$D(P_0^\pi || P_1^\pi)$

- Suppose policy π says "no matter what, stop at T".
- $x = (a_1, x_1, a_2, x_2, \ldots, a_{T-1}, x_{T-1}, a_T = stop)$.
- 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)

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

$D(Q_0^\pi || Q_1^\pi)$, and a summing up

- Interpretation of Q_0^π: Under hypothesis H_0, when you stop, probabilities of various decisions

-

Hypothesis	Distribution	Decision 0	Decision 1
H_0	Q_0^π	$\geq 1-\varepsilon$	$\leq \varepsilon$
H_1	Q_1^π	$\leq \varepsilon$	$\geq 1-\varepsilon$

- Approximately $D(\{1-\varepsilon, \varepsilon\}||\{\varepsilon, 1-\varepsilon\})$

$$(1-\varepsilon)\log\frac{1-\varepsilon}{\varepsilon} + \varepsilon\log\frac{\varepsilon}{1-\varepsilon} \sim \log\frac{1}{\varepsilon}.$$

- Thus: $E_0[\tau]D(p_0||p_1) \gtrsim \log\frac{1}{\varepsilon}$, or

$$E_0[\tau] \gtrsim \frac{\log\left(\frac{1}{\varepsilon}\right)}{D(p_0||p_1)}.$$

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 $\frac{P_0^\pi(\cdots)}{P_1^\pi(\cdots)} \geq \frac{1}{\varepsilon}$.
  This is what makes it an ε-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

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

$N =$ 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 (λ_0, λ_1) pairs.
- You should try:

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

  $m = 24, \Delta = 250$ *ms* from the Sripati and Olson experiments.

Assignment: Correlation analysis

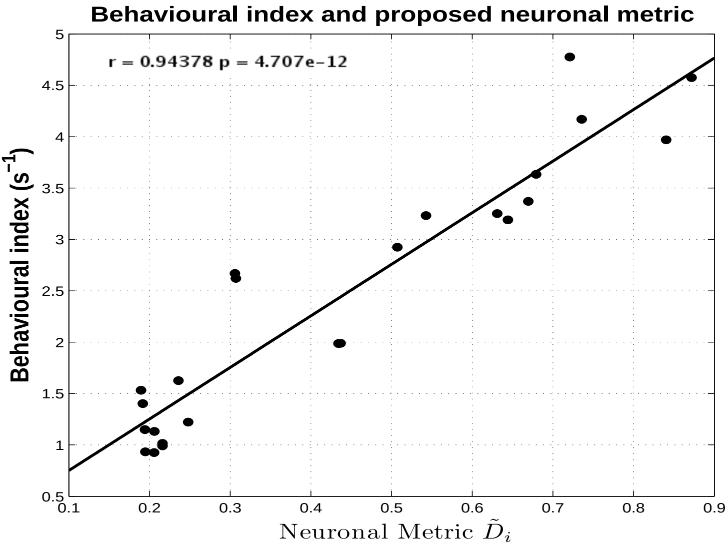
- Divide data into groups. Each group is for an ordered image pair.

- Compute s_{ij}, $\hat{D}_{ij}$, L_{ij}.
  s_{ij} plays the role of τ.
  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



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 $g(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x \geq 0$.
  α is the shape, β is the rate.
- Mean $= \alpha/\beta$, stddev $= \sqrt{\alpha}/\beta$, so that stddev/mean $= 1/\sqrt{\alpha}$.
  Fit a straight line to the scatter plot above and provide a guess for the shape α.

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.

$$
\begin{aligned}
D(P_h^\pi || P_{h'}^\pi) &= E_h^\pi \sum_{t=1}^{\tau} \log \frac{p_h^{A_t}(X_t)}{p_{h'}^{A_t}(X_t)} \quad \text{(conditional independence)} \\
&= E_h^\pi \sum_{a=1}^{K} \sum_{l=1}^{N_a(\tau)} \log \frac{p_h^a(X_l)}{p_{h'}^a(X_l)} \\
&= \sum_{a=1}^{K} E_h^\pi [N_a(\tau)] D(p_h^a || p_{h'}^a) \quad \text{(Optional stopping)} \\
&\leq E_h^\pi [\tau] \quad \max_{\lambda} \sum_{a=1}^{K} \lambda_a D(p_h^a || p_{h'}^a).
\end{aligned}
$$

- How should an adversary choose h' to minimise the information content in each slot? How should the searcher choose λ 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.