# Bayesian Causal Inference in Perception

March 29, 2018

- Motivation: what is a problem the brain must solve?
  - avoid/minimize redundancy and uncertainty
  - combine/optimize useful information
- what are some real world instances ~~that~~ examples situations in which the brain must do the above?
  - sensorimotor integration
    - ↳ Kalman filters
  - object perception
    - ↳ visual illusions!
  - multisensory integration
    - ↳ speech perception/McGurk

  "optimal ~~was~~ ways to combine info to estimate something about the world

**Today**: Multisensory Integration

- world is noisy, senses inaccurate
- how can brain combine redundant information?

**Idea 1**: Ernst & Banks 2002 | Maximum Likelihood Estimation

vision  $S_V$ touch  $S_H$ judging the height of a raised ridge

MLE: $\hat{S} = \sum_i w_i \hat{s}_i$ $\qquad w_i = \dfrac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$

optimal estimate assuming:
- noises are independent across modalities and are Gaussian with variance $\sigma_i^2$

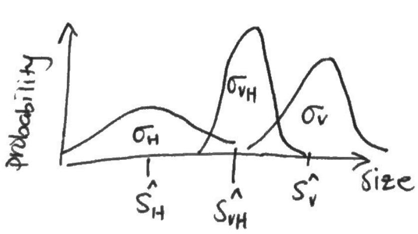
... is to add estimates weighted by normalized reciprocal variances

$$W_V = \frac{\frac{1}{\sigma_V^2}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_H^2}} \quad \text{and} \quad W_H = \frac{\frac{1}{\sigma_H^2}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_H^2}}, \quad W_V + W_H = 1$$

$$\boxed{\hat{S}_{VH} = \frac{\frac{\hat{S}_V}{\sigma_V^2}+\frac{\hat{S}_H}{\sigma_H^2}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_H^2}}}$$

- experimentally, researchers measured unisensory discrimination thresholds $\Rightarrow \frac{1}{\sigma_V^2} \quad \frac{1}{\sigma_H^2}$ and percieved unisensory heights $\Rightarrow \hat{S}_V \quad \hat{S}_H$
- For bimodal condition, $\hat{S}_{VH}$, MLE predicts that combined estimate has lower variance (lower discrimination thresholds) than either alone

* intuition *




product of two Gaussians ⇒ variance decreases

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \Rightarrow \frac{\sigma_H^2}{\sigma_V^2\sigma_H^2} + \frac{\sigma_V^2}{\sigma_V^2\sigma_H^2}$$

$$\Rightarrow \frac{\sigma_H^2+\sigma_V^2}{\sigma_V^2\sigma_H^2} = \frac{1}{\sigma_{VH}^2}$$

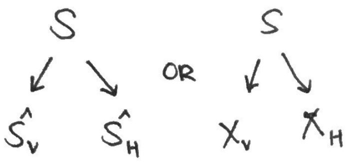
- Experimentally, found that behavior matches MLE predictions: multisensory estimate is a weighted ~~linear average~~ of unisensory cues

$$\boxed{\sigma_{VH}^2 = \frac{\sigma_V^2\sigma_H^2}{\sigma_H^2+\sigma_V^2}}$$

## Idea 2: Kording et. al. 2007 | Causal Inference Perspective

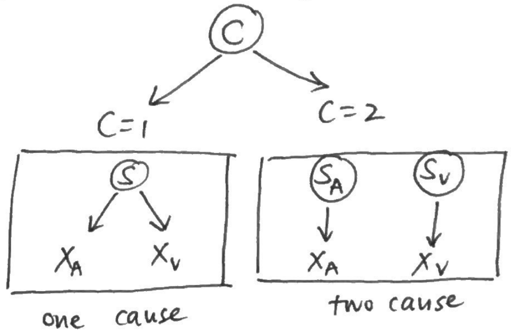
- instead of only taking sensory signals into account, maybe brain must also take into account where signals came from (cause) $S \rightarrow X$

**before:**




(we will call them X now to specify that they are internal variables)

**now:** consider the generative model




task: report location of visual flash + auditory cue

Stimulus

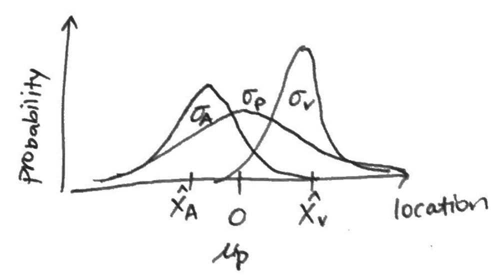
subjects are biased by cues!! either think that two cues come from same location (C=1) or different locations (C=2)

causal inference ingredients:

likelihood: $p(X_A|S_A)$ $p(X_V|S_V)$

prior: how likely two cooccuring signals have 1 vs 2 causes

$p(C)$

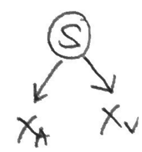



**Step 1:** estimate probability that two cues came from same cause

$$P(C|X_V,X_A) = \frac{P(X_V, X_A|C)\,P(C)}{P(X_V, X_A)}$$

one note: $P(C=1|X_V,X_A) + P(C=2|X_V,X_A) = 1$

$$P(C=1|X_V,X_A) = \frac{P(X_V,X_A|C=1)\,P(C=1)}{P(X_V,X_A|C=1)\,P(C=1) + P(X_V,X_A|C=2)(1-P(C=1))}$$




breaking it down... $P(X_V,X_A|C=1) = \int p(X_V,X_A|S)\,p(S)\,dS$

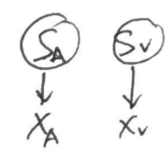
because sensory noise is independent...

$$= \int p(X_V|S)\,p(X_A|S)\,p(S)\,dS$$

all 3 factors here are Gaussian... analytic solution

$$p(X_V,X_A|C=1) = \frac{1}{2\pi\sqrt{\sigma_V^2\sigma_A^2+\sigma_V^2\sigma_p^2+\sigma_A^2\sigma_p^2}}\, e^{-\frac{1}{2}\frac{(X_V-X_A)^2\sigma_p^2+(X_V-\mu_p)^2\sigma_A^2+(X_A-\mu_p)^2\sigma_V^2}{\sigma_V^2\sigma_A^2+\sigma_V^2\sigma_p^2+\sigma_A\sigma_p^2}}$$

For $p(X_V, X_A | C=2)$, $X_V$ and $X_A$ are independent, we obtain a product of 2 factors:




$$p(X_V, X_A | C=2) = \cancel{\iint p(X_V|S_V)p(S_V)dS_V}$$

$$\iint p(X_V, X_A | S_V, S_A)\, p(S_V, S_A)\, dS_V\, dS_A =$$

$$\left(\int p(X_V|S_V)p(S_V)dS_V\right)\left(\int p(X_A|S_A)p(S_A)dS_A\right)$$

$$p(X_V, X_A | C=2) = \frac{1}{2\pi\sqrt{(\sigma_V^2+\sigma_p^2)(\sigma_A^2+\sigma_p^2)}}\, e^{-\frac{1}{2}\left[\frac{(X_V-\mu_p)^2}{\sigma_V^2+\sigma_p^2} + \frac{(X_A-\mu_p)^2}{\sigma_A^2+\sigma_p^2}\right]}$$

Step 2: estimate location of auditory and visual cues given $p(C|X_A, X_V)$

$P(\hat{S_V}|X_V, X_A)$ or $P(\hat{S_A}|X_V, X_A)$

- like in part 1, each estimate given C=1 OR C=2 is a weighted linear average of unisensory cues
- under C=1:

$$\hat{S}_{V,C=1} = \hat{S}_{A,C=1} = \frac{\frac{X_V}{\sigma_V^2} + \frac{X_A}{\sigma_A^2} + \frac{\mu_p}{\sigma_p^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_A^2} + \frac{1}{\sigma_p^2}}$$

- under C=2: optimal estimate would be same as when there would only be a visual or only an auditory signal

$$\hat{S}_{V,C=2} = \frac{\frac{X_V}{\sigma_V^2} + \frac{\mu_p}{\sigma_p^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_p^2}} \qquad \hat{S}_{A,C=2} = \frac{\frac{X_A}{\sigma_A^2} + \frac{\mu_p}{\sigma_p^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_p^2}}$$

Step 3:

Finally, combine (weight) the percepts under C=1 and C=2 by the probability of each causal structure:

$$\hat{S_V} = p(C=1|X_V, X_A)\cdot \hat{S}_{V,C=1} + p(C=2|X_V, X_A)\cdot \hat{S}_{V,C=2}$$

$$\hat{S_A} = p(C=1|X_V, X_A)\cdot \hat{S}_{A,C=1} + p(C=2|X_V, X_A)\cdot \hat{S}_{A,C=2}$$