Signal Detection Theory

Ekaterina Kravtchenko
Ling 239 / March 13, 2012

11.1 Signal Detection Theory

Signal Detection Theory is used (primarily) in psychology for making inferences from data involving decision-making in the light of uncertainty.

It's mostly applied for two-alternative forced-choice experiments where participants indicate whether they think that they [don't] perceive a given stimulus, but technically can be used for any 2x2 table of counts.

Data/Terminology

Signal Trials: Signal to be detected present.

Noise Trials: Signal not present.

Yes/No Responses: Whether the 'signal' is [in]correctly detected by the participant.

- **Hit:** [Correct] 'Yes' response to Signal trial.
- **False Alarm:** [Incorrect] 'Yes' response to Noise trial
- **Miss:** [Incorrect] 'No' response to Signal trial.
- **Correct Rejection:** [Correct] 'No' response to Noise trial

	Signal Trial	Noise Trial
Yes Response	Hit	False Alarm
No Response	Miss	Correct Rejection

Basic Data: Counts of hits/FAs/misses/CRs.

Also Common: If consider just hit and FA counts, together with # of Signal/Noise trials, describe the entirety of the data

SDT Assumptions

Key Assumptions: Representation and Decision-Making

Representation: Signal/Noise trials can be represented as values that correspond to a uni-dimensional measure of 'strength'; both types have 'strengths' that vary according to a Gaussian distribution

Signal Strength: Assumed to be, on average, greater than Noise strengths, so this distribution has a higher mean.

Equal-Variance SDT: Most common; distributions assumed to have same variance.

Decision-Making: Comparing strength of current trials to some fixed criterion; this produces Yes/No responses. If strength exceeds criterion $\rightarrow$ Yes response; otherwise No response.

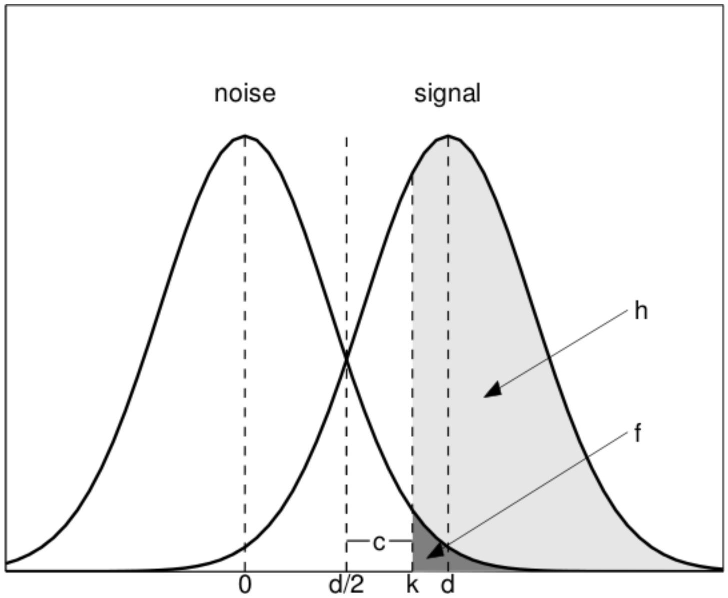
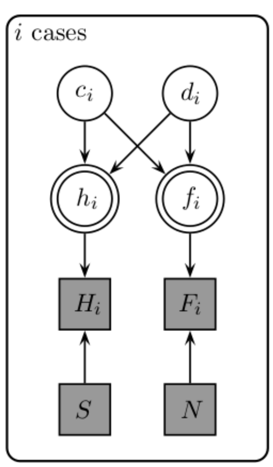


Figure 1: Equal-variance Gaussian SDT Framework

The above is a formal version of the equal-variance SDT model.

- Arbitrary units of underlying strength scale → variances fixed to 1 and mean of noise distribution set to 0.
- **d: mean of signal distribution** → b/c corresponds to distance b/w two distributions, measure of discriminability of Signal/Nose trials from each other.
- **d/2**: strength value at which Signal/Noise distributions equally likely → corresponds to unbiased responding.
- **k**: criterion used for responding.
- **c**: 'unbiased criterion' (how different actual criterion is from unbiased criterion)
  - → Measure of bias.
  - Positive values: bias towards No response. (More CRs & Misses)
  - Negative values: bias towards Yes response. (More Hits & FAs)



$$
\begin{aligned}
d_i &\sim \text{Gaussian}(0, \tfrac{1}{2}) \\
c_i &\sim \text{Gaussian}(0, 2) \\
h_i &\leftarrow \Phi(\tfrac{1}{2}d_i - c_i) \\
f_i &\leftarrow \Phi(-\tfrac{1}{2}d_i - c_i) \\
H_i &\sim \text{Binomial}(h_i, S) \\
F_i &\sim \text{Binomial}(f_i, N)
\end{aligned}
$$

Figure 2: Graphical model for SDT

The above is a graphical model for inferring **Discriminability** & **Bias** from Hit (h_i)/FA (f_i) counts for the ith case.

The Hit/FA rates are functions of their associated discriminabilities d_i and biases c_i (using cumulative standard Normal distribution function $\Phi(\cdot)$.

The observed counts of Hits (H_i) & FA rates (F_i) are Binomially distributed according to hits/FA rates & # of Signal trials (S) & Noise trials (N).

Discriminability & Bias priors are both Gaussian distributions – correspond to uniform prior distribution over Hit/FA rates.

Predictions about Hit/FA Rates

1. **h**: proportion of signal distribution above criterion **k**
2. **FA rate (f)** = proportion of noise distribution above **k**

SDT can therefore take data such as that in the table on the first page, and convert Hit/FA counts into measures of discriminability/bias.

Discriminability: How easily Signal/Noise trials distinguished.

Bias: How decision-making criterion relates to optimal criterion.

	Control Group		Group I		Group II	
	Old Odor	New Odor	Old Odor	New Odor	Old Odor	New Odor
Old Resp.	148	29	150	40	150	51
New Resp.	32	151	30	140	40	139

Table 1: Recognition Memory for Odors – Lehrner et al. (1995)

Exercise: ...Analyze these three data sets using signal detection theory to infer the discriminability and bias of the recognition performance for each group. What conclusions do you draw from this analysis? What, if anything, can you infer about individual differences between the subjects in the same groups?

11.2 Hierarchical Signal Detection Theory

Individual subject data available $\rightarrow$ model individual differences using hierarchical extension of SDT

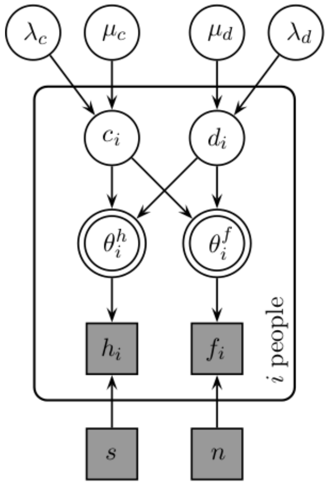
Idea: Diff. subjects have different discriminabilities & biases drawn from group-level Gaussian distributions.

Framework can be used for looking at e.g. inductive/deductive reasoning (Heit & Rotello, 2005).

- **Argument Strength** is uni-dimensional construct – used as signal strength.
- Different criteria allowed for induction/deduction (**Induction**: weak/strong arguments; **Deduction**: valid/invalid arguments).
- Deductive criterion is more extreme $\rightarrow$ **Deduction = more stringent form of Induction**.
- SDT modeling of these phenomena has implications for debate over diff. kinds of reasoning systems/processes.

Heit and Rotello (2005):

- Inductive/Deductive judgments of 80 participants on 8 arguments.
- 40 subjects asked Induction Qs (if conclusion "plausible"); 40 Deduction Qs (if conclusion "necessarily true") $\rightarrow$ can be characterized by Hit/FA counts.



$$
\begin{aligned}
\mu_d, \mu_c &\sim \text{Gaussian}(0, .001)\\
\lambda_d, \lambda_c &\sim \text{Gamma}(.001, .001)\\
d_i &\sim \text{Gaussian}(\mu_d, \lambda_d)\\
c_i &\sim \text{Gaussian}(\mu_c, \lambda_c)\\
\theta_i^h &\leftarrow \Phi(\tfrac{1}{2}d_i - c_i)\\
\theta_i^f &\leftarrow \Phi(-\tfrac{1}{2}d_i - c_i)\\
h_i &\sim \text{Binomial}(\theta_i^h, s)\\
f_i &\sim \text{Binomial}(\theta_i^f, n)
\end{aligned}
$$

Figure 3: Graphical model for Hierarchical SDT

The above uses SDT to infer **Discriminability** d_i and **Bias** c_i from Hit (H_i) & FA (F_i) counts, for the *i*th subject.

Individual differences modeled hierarchically – assumption that individual discriminabilities & biases come from Gaussian group-level distributions. **Means**: μ_d, μ_c; **Precisions**: λ_d, λ_c.

Joint posterior over group means for **Discriminability** & **Bias**, for both experimental conditions:

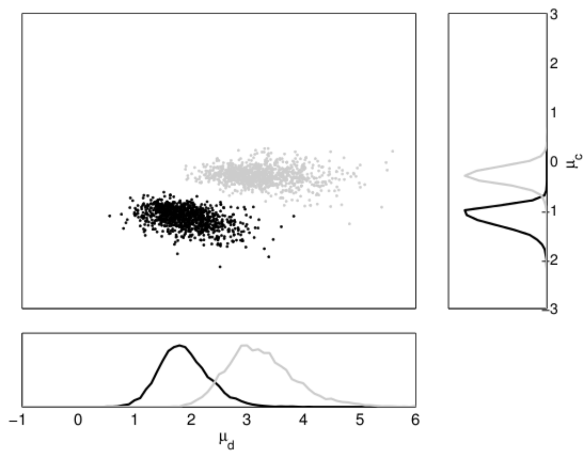


Figure 4: Joint posterior over μ_d and μ_c for Induction (dark) and Deduction (light) conditions