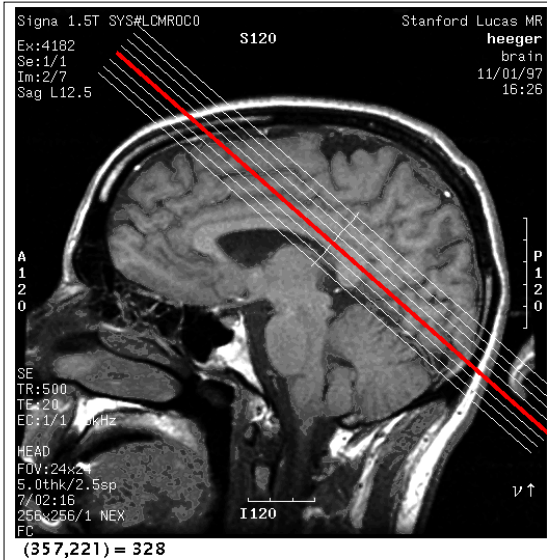


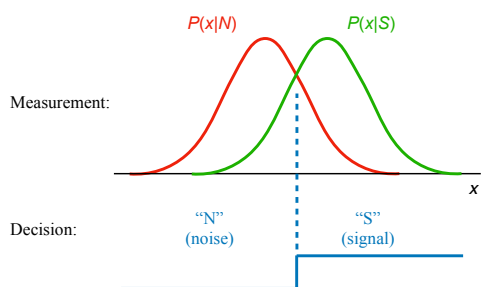
Mathematical Tools for Neural and Cognitive Science

Fall semester, 2023

Section 5a: Statistical Decision Theory + Signal Detection Theory

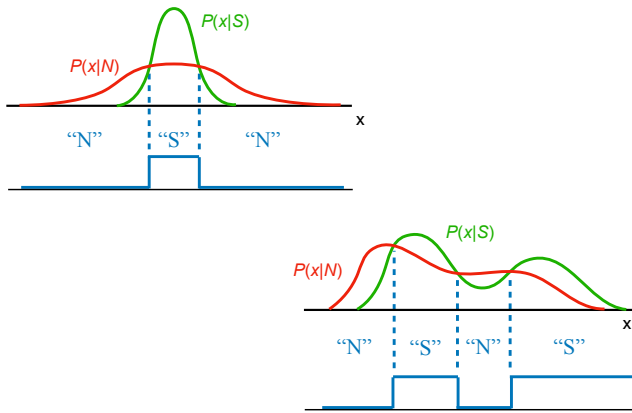


Signal Detection Theory (binary estimation)

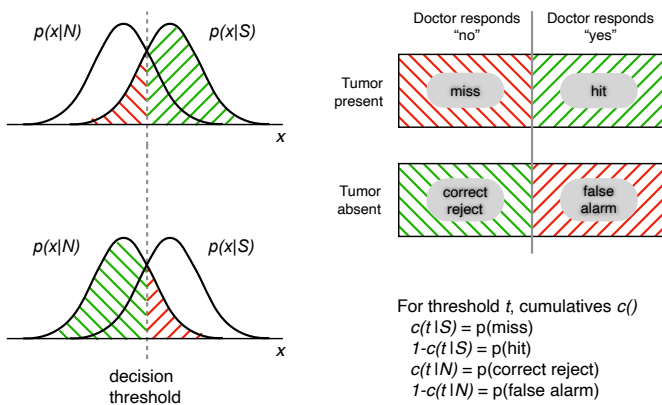


For equal-shape, unimodal, symmetric distributions, the ML decision rule is a *threshold* function.

More generally, decision rule can have multiple thresholds...

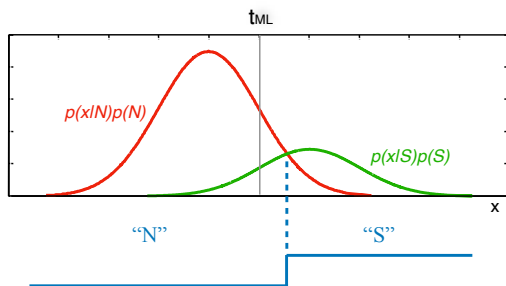


Signal Detection Theory: Potential outcomes



MAP decision rule?

MAP solution maximizes proportion of correct answers, *taking prior probability into account*.



Compared to ML threshold, the MAP threshold moves *away* from higher-probability option.

Bayes decision rule?

Incorporate values for the four possible outcomes:

Payoff Matrix

		Response	
		Yes	No
Stimulus	S+N	V_{S+N}^{Yes}	V_{S+N}^{No}
	N	V_N^{Yes}	V_N^{No}

Bayes Optimal Criterion

		Response	
		Yes	No
Stimulus	S+N	V_{S+N}^{Yes}	V_{S+N}^{No}
	N	V_N^{Yes}	V_N^{No}

$$\mathbb{E}(Yes|x) = V_{S+N}^{Yes}p(S+N|x) + V_N^{Yes}p(N|x)$$

$$\mathbb{E}(No|x) = V_{S+N}^{No}p(S+N|x) + V_N^{No}p(N|x)$$

Say yes if $\mathbb{E}(Yes|x) \geq \mathbb{E}(No|x)$

Optimal Criterion

$$\mathbb{E}(Yes|x) = V_{S+N}^{Yes}p(S+N|x) + V_N^{Yes}p(N|x)$$

$$\mathbb{E}(No|x) = V_{S+N}^{No}p(S+N|x) + V_N^{No}p(N|x)$$

Say yes if $\mathbb{E}(Yes|x) \geq \mathbb{E}(No|x)$

$$\text{Say yes if } \frac{p(S+N|x)}{p(N|x)} \geq \frac{V_N^{No} - V_N^{Yes}}{V_{S+N}^{Yes} - V_{S+N}^{No}} = \frac{V(\text{Correct}|N)}{V(\text{Correct}|S+N)}$$

Posterior odds

Apply Bayes' Rule

$$p(S+N | x) = \frac{p(x | S+N)p(S+N)}{p(x)}$$

Posterior
Likelihood
Prior
Nuisance normalizing term

$$p(N | x) = \frac{p(x | N)p(N)}{p(x)}, \text{ hence}$$

$$\frac{p(S+N | x)}{p(N | x)} = \left(\frac{p(x | S+N)}{p(x | N)} \right) \left(\frac{p(S+N)}{p(N)} \right)$$

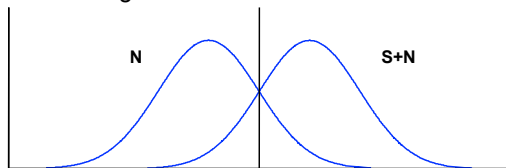
Posterior odds
Likelihood ratio
Prior odds

Optimal Criterion

Say yes if $\frac{p(S+N | x)}{p(N | x)} \geq \frac{V(\text{Correct} | N)}{V(\text{Correct} | S+N)}$

i.e., if $\frac{p(x | S+N)}{p(x | N)} \geq \frac{p(N)}{p(S+N)} \frac{V(\text{Correct} | N)}{V(\text{Correct} | S+N)} = \beta$

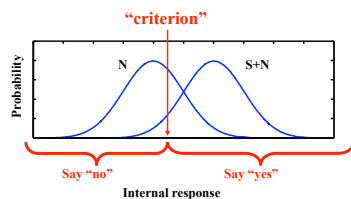
Example, if equal priors and equal payoffs, say yes if the likelihood ratio is greater than one:



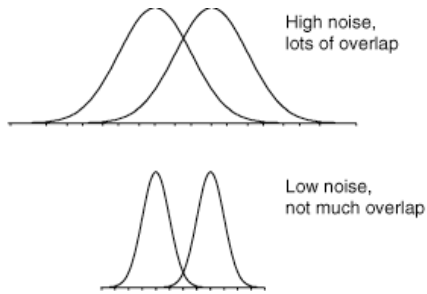
Example applications of SDT

- Vision
 - Detection (something vs. nothing)
 - Discrimination (lower vs greater level of: intensity, contrast, depth, slant, size, frequency, loudness, ...)
- Memory (internal response = trace strength = familiarity)
- Neurometric function/discrimination by neurons (internal response = spike count)

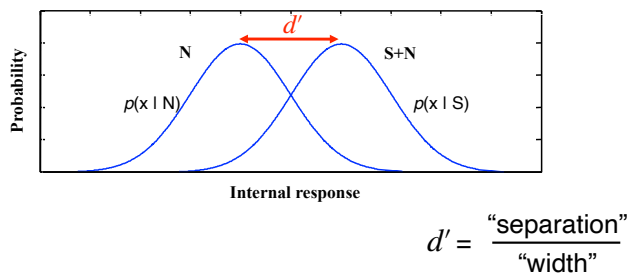
From experimental measurements, assuming human is optimal, can we determine the underlying distributions and criterion?



Signal Detection Theory: discriminability (d')



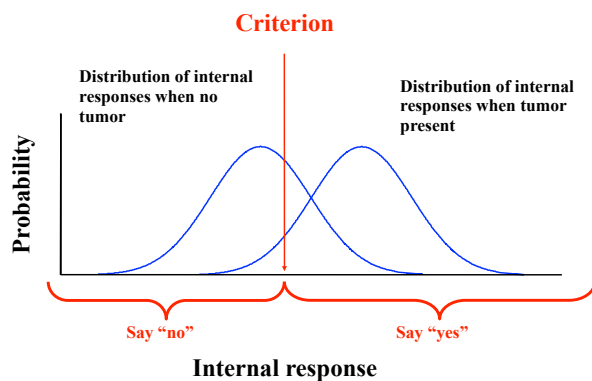
Internal response: probability of occurrence curves



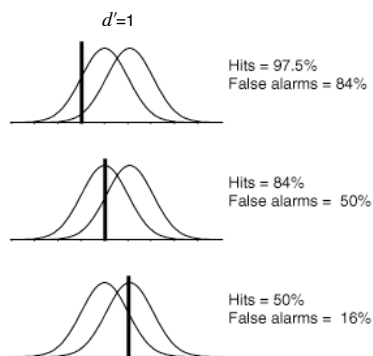
Discriminability ("d-prime") is the normalized separation between the two distributions

Error rate is a function of d'

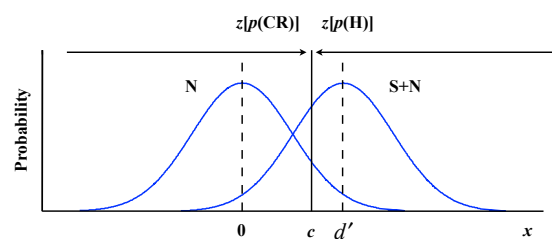
Criterion



Signal Detection Theory: Criterion



SDT: Gaussian case



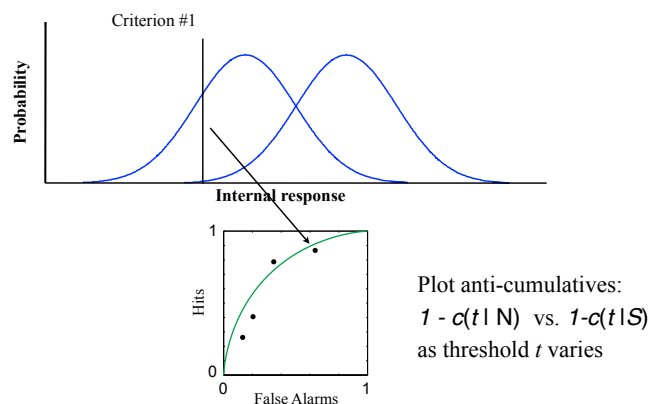
$$d' = z[p(H)] + z[p(CR)] = z[p(H)] - z[p(FA)]$$

$$c = z[p(CR)]$$

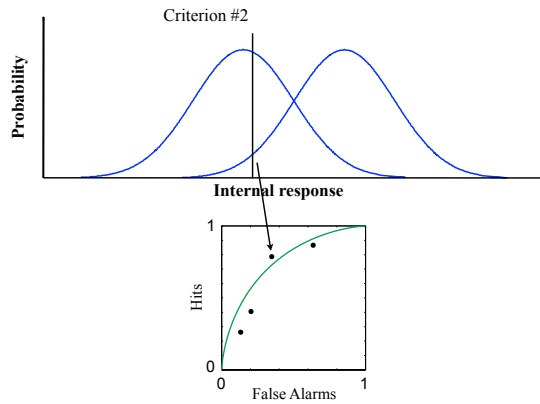
$$G(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$\beta = \frac{p(x=c | S+N)}{p(x=c | N)} = \frac{e^{-(c-d')^2/2}}{e^{-c^2/2}} \quad (\text{Fix } \sigma = 1)$$

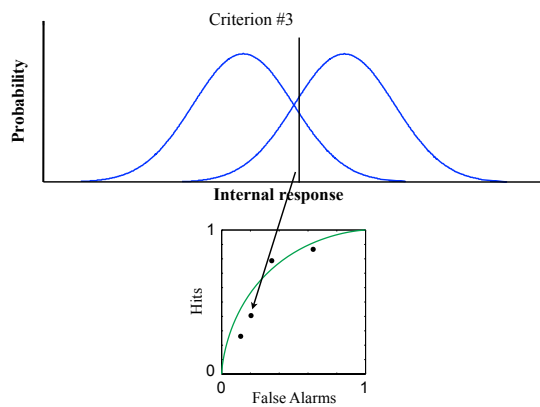
ROC (Receiver Operating Characteristic)



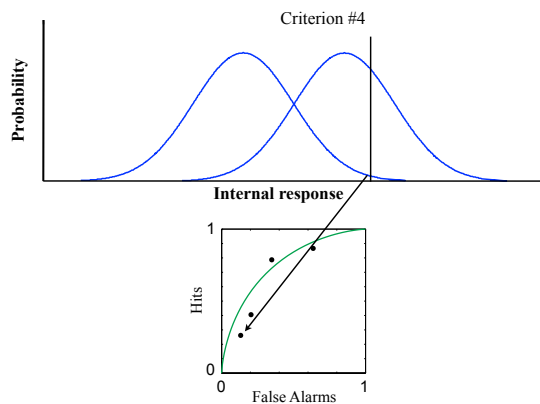
ROC (Receiver Operating Characteristic)



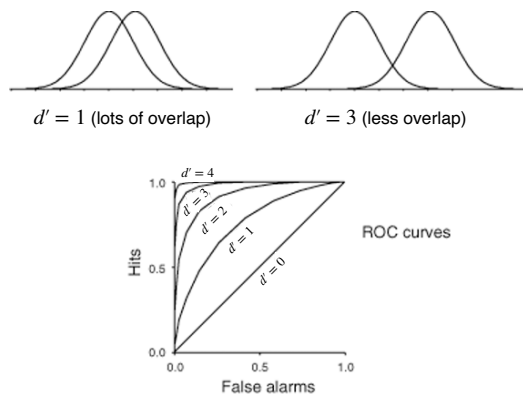
ROC (Receiver Operating Characteristic)



ROC (Receiver Operating Characteristic)



ROC (Receiver Operating Characteristic)



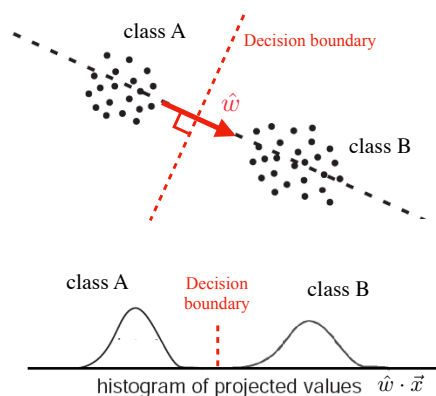
[on board: Area under curve = %correct in a 2AFC task]

Decision/classification in multiple dimensions

- Data-driven linear classifiers:
 - Prototype Classifier - minimize distance to class mean
 - Fisher Linear Discriminant (FLD) - maximize d'
 - Support Vector Machine (SVM) - maximize margin
- Statistical:
 - ML/MAP/Bayes under a probabilistic model
 - e.g.: Gaussian, identity covariance (same as Prototype)
 - e.g.: Gaussian, equal covariance (same as FLD)
 - e.g.: Gaussian, general case (Quadratic Discriminator)
- Some Examples:
 - Visual gender classification
 - Neural population decoding

Linear Classifier

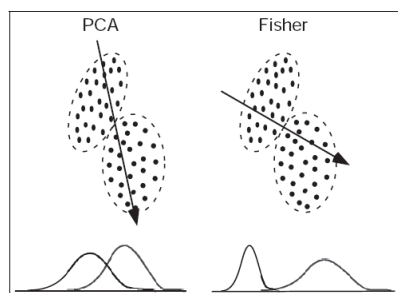
Find unit vector \hat{w} ("discriminant") that best separates the distributions



Simplest linear discriminant: the Prototype Classifier

$$\hat{w} = \frac{\vec{\mu}_A - \vec{\mu}_B}{\|\vec{\mu}_A - \vec{\mu}_B\|}$$

Fisher Linear Discriminant



$$\max_{\hat{w}} \frac{[\hat{w}^T(\vec{u}_A - \vec{u}_B)]^2}{[\hat{w}^T C_A \hat{w} + \hat{w}^T C_B \hat{w}]} \quad (\text{note: this is } d'^2 \text{ squared!})$$

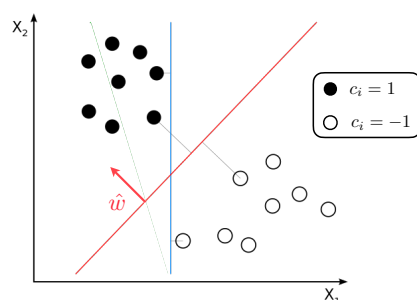
$$\text{optimum: } \hat{w} = C^{-1}(\vec{u}_A - \vec{u}_B), \text{ where } C = \frac{1}{2}(C_A + C_B)$$

Support Vector Machine (SVM)

(widely used in machine learning, but no closed form solution)

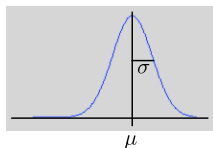
Maximize the “margin” (gap between data sets):

find largest m , and $\{\hat{w}, b\}$ s.t. $c_i(\hat{w}^T \vec{x}_i - b) \geq m, \quad \forall i$

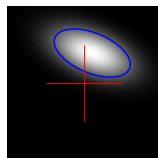


Reminder: Multi-D Gaussian densities

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



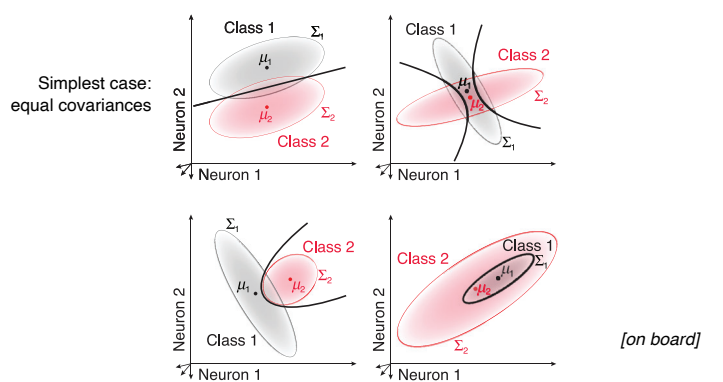
$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |C|}} e^{-\frac{(\vec{x}-\vec{\mu})^T C^{-1} (\vec{x}-\vec{\mu})}{2}}$$



mean: [0.2, 0.8]
cov: [1.0 -0.3;
-0.3 0.4]

ML (or MAP) classifier for two Gaussians

Decision boundary is *quadratic*, with four possible geometries:



[figure: Pagan et al. 2016]

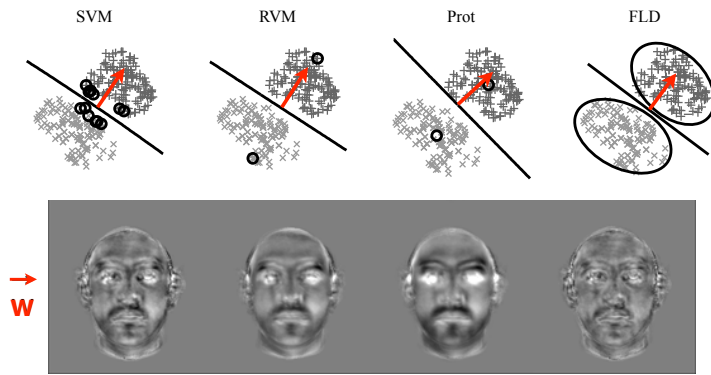
A perceptual example: Gender identification



- 200 face images (100 male, 100 female)
- Adjusted for position, size, intensity/contrast
- Labeled by 27 human subjects

[Graf & Wichmann, NIPS*03]

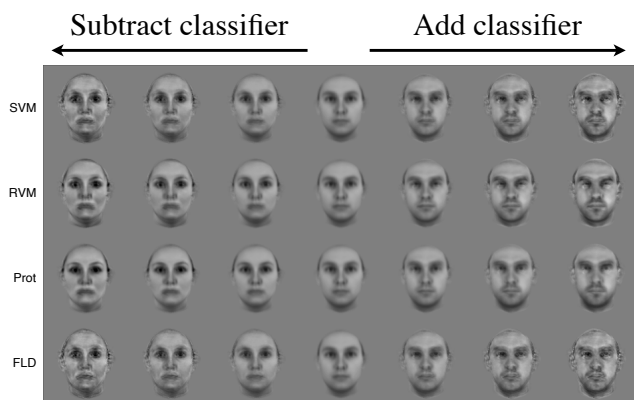
Linear classifiers



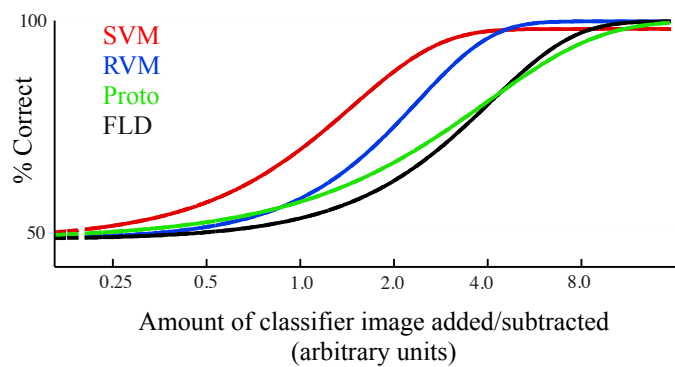
Four linear classifiers, trained on human data

Model validation/testing

- Cross-validation: Subject responses [% correct, reaction time, confidence] are explained
 - very well by SVM
 - moderately well by RVM / FLD
 - not so well by Prot
- Do these decision “models” make testable predictions? Synthesize optimally discriminable faces...



[Wichmann, et. al; NIPS*04]



[Wichmann, et. al; NIPS*04]

Fisher Information

- Second-order expansion of the (expected) negative log likelihood:

$$I(s) = -\mathbb{E} \left[\frac{\partial^2 \log p(r|s)}{\partial s^2} \right]$$

- Provides a bound on “precision” of unbiased estimators: (the “Cramér-Rao bound”) $\sigma^2(s) \geq \frac{1}{I(s)}$

- Perceptually, provides a bound on **discriminability**: (Series et. al. 2009) $D(s) \leq \sqrt{I(s)}$

- Examples: with mean stimulus response $\mu(s)$

Gaussian case: $p(r|s) \sim \mathcal{N}(\mu(s), \sigma^2)$ $I(s) = [\mu'(s)]^2 / \sigma^2$

Poisson case: $p(r|s) \sim \text{Poiss}(\mu(s))$ $I(s) = [\mu'(s)]^2 / \mu(s)$

Example: Weber’s law

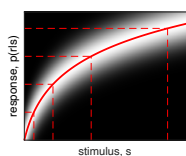
[Weber, 1834]

$$D(s) \propto \frac{1}{s} \quad (\text{discrimination thresholds proportional to stimulus strength})$$

Assuming $I(s) \propto \frac{1}{s^2}$ what internal representation can explain this? Many!

additive Gaussian noise, with mean

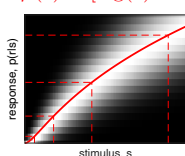
$$\mu(s) = \log(s) + c$$



entirely due to response mean
[Fechner, 1860]

Poisson noise, with mean

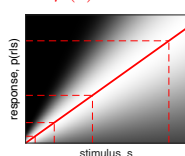
$$\mu(s) = [\log(s) + c]^2$$



discrete representation, depends on both mean and variance

multiplicative Gaussian noise, with mean

$$\mu(s) = \alpha s$$



entirely due to response variance

S.S. Stevens. “To Honor Fechner and Repeal His Law: A power function, not a log function, describes the operating characteristic of a sensory system” (1961)

Three examples with different power-law mean response, each consistent with Weber’s law discriminability.

