MODVIS 2025

Statistical Decision Theory
+
Signal Detection Theory

Michael S. Landy
Dept. of Psychology & Center for Neural Science
NYU

Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

Landy, M. S. (2025/in press). Signal detection theory. In Nichols, A. L. & Edlund, J. (Eds.), *Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences*, Volume 3, Chapter 24. New York: Cambridge University Press.

References

Books:

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics. Krieger.*

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge.

Wickens, T. D. (2002). *Elementary Signal Detection Theory.* Oxford University Press.

Classic papers:

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions IRE Professional Group on Information Theory*, PGIT-4, 171-212.

Van Meter, D., & Middleton, D. (1954). Modern statistical approaches to reception in communication theory. *Transactions IRE Professional Group on Information Theory*, *PGIT-4*, 119-145.

Slides

Paper

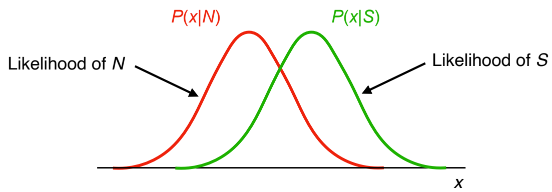
Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

(357,221) = 328

Tumor, or not?

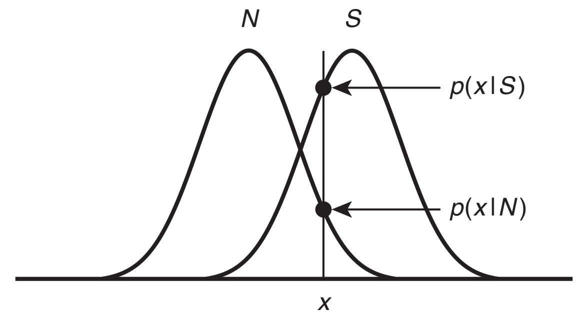
Signal Detection Theory



Before the trial starts $P(x \mid S)$ and $P(x \mid N)$ are functions of x, the "measurement distribution".

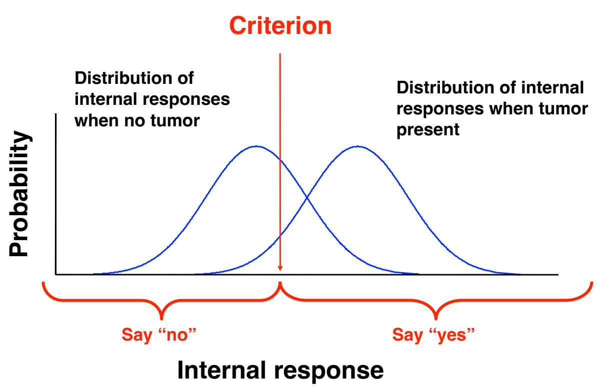
After the stimulus is presented, from the observer's perspective $P(x \mid S)$ and $P(x \mid N)$ are functions of the second variable (S or N) and are now called the "likelihood function".

Signal Detection Theory

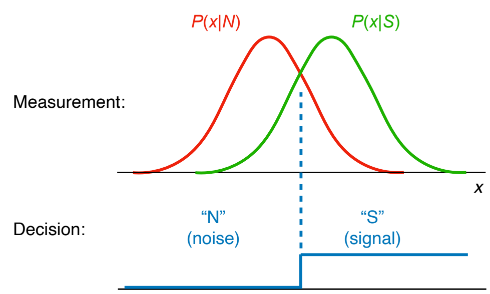


After stimulus presentation, the observer has measurement x and thus can infer the likelihoods $P(x \mid N)$ and $P(x \mid S)$ and use their values to inform a decision.

Criterion

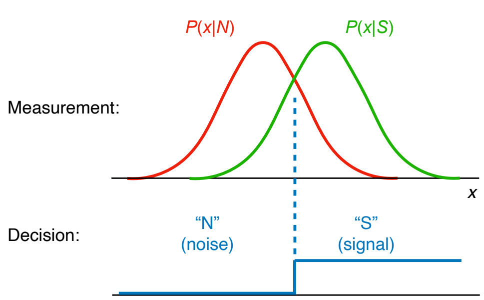


Signal Detection Theory



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

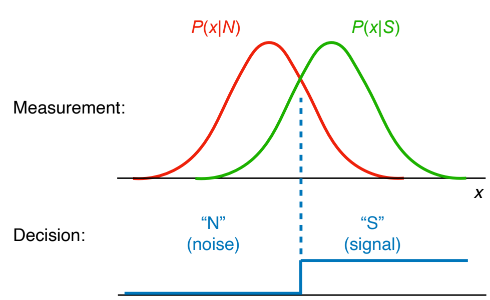
Signal Detection Theory



Say "yes" if $p(x \mid S) > p(x \mid N)$

Say "no" otherwise.

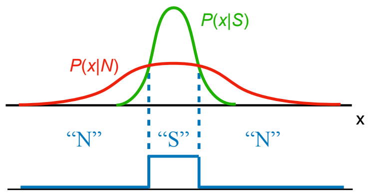
Signal Detection Theory

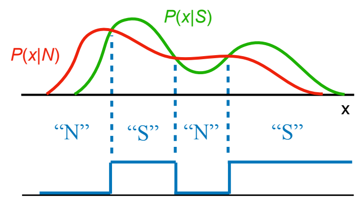


Say "yes" if $x > \dfrac{\mu_S + \mu_N}{2} = c$

Say "no" otherwise.

More generally, an ML decision rule can have multiple thresholds:





Decision rules

Maximum likelihood (ML):

Say “yes” if $p(x\,|\,S) > p(x\,|\,N)$
Say “no” otherwise.

Maximum *a posteriori* (MAP):

Say “yes” if $p(S\,|\,x) > p(N\,|\,x)$
Say “no” otherwise.

Apply Bayes’ Rule

$$\underbrace{P(S\,|\,x)}_{\text{Posterior}} = \frac{\overbrace{p(x\,|\,S)}^{\text{Likelihood}}\overbrace{P(S)}^{\text{Prior}}}{\underbrace{p(x)}_{\text{Nuisance normalizing term}}}$$

Decision rules

Maximum *a posteriori* (MAP):

Say “yes” if $p(S\,|\,x) > p(N\,|\,x)$
Say “no” otherwise.

Decision rules

Maximum *a posteriori* (MAP):

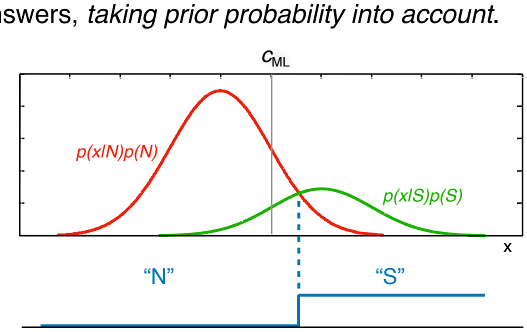
Say “yes” if $p(S\,|\,x) > p(N\,|\,x)$
Say “no” otherwise.

Say “yes” if $\dfrac{p(x\,|\,S)P(S)}{p(x)} > \dfrac{p(x\,|\,N)P(N)}{p(x)}$
Say “no” otherwise.

Say “yes” if $p(x\,|\,S)P(S) > p(x\,|\,N)P(N)$
Say “no” otherwise.

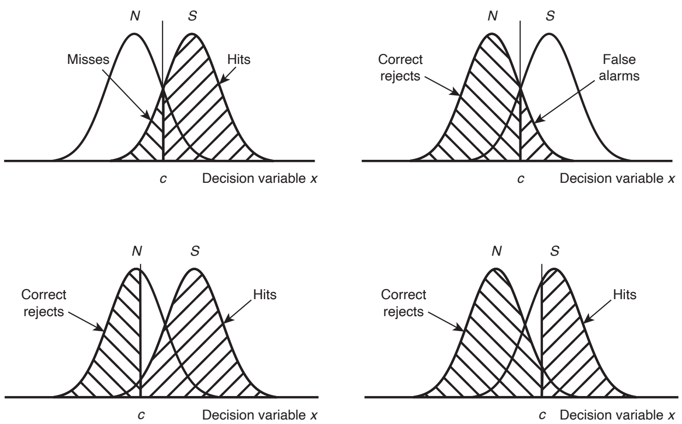
MAP decision rule

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

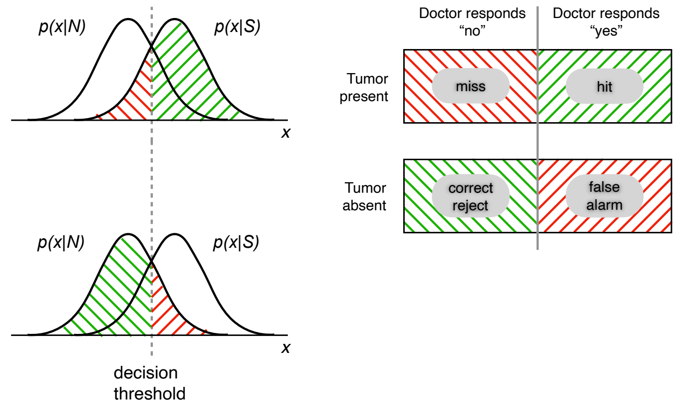


Compared to ML threshold, the MAP threshold moves *away* from higher-probability option.

Signal Detection Theory: Potential outcomes



Signal Detection Theory: Potential outcomes



Bayes decision rule (maximum expected gain or minimum Bayes risk)

Incorporate values for the four possible outcomes:

Payoff Matrix

Stimulus	Response: No	Response: Yes
S	V_S^{No}	V_S^{Yes}
N	V_N^{No}	V_N^{Yes}

Bayes Optimal Criterion

Stimulus	Response: No	Response: Yes
S	V_S^{No}	V_S^{Yes}
N	V_N^{No}	V_N^{Yes}

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

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

Optimal Criterion

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

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

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

Posterior odds

Apply Bayes' Rule

$$P(S\,|\,x) = \frac{p(x\,|\,S)P(S)}{p(x)}$$

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

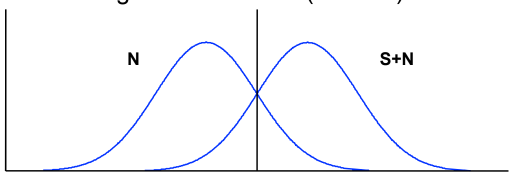
$$\underbrace{\frac{P(S\,|\,x)}{P(N\,|\,x)}}_{\text{Posterior odds}} = \underbrace{\left(\frac{p(x\,|\,S)}{p(x\,|\,N)}\right)}_{\text{Likelihood ratio}} \underbrace{\left(\frac{P(S)}{P(N)}\right)}_{\text{Prior odds}}$$

Optimal Criterion

Say yes if $\dfrac{P(S\,|\,x)}{P(N\,|\,x} \geq \dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}$

i.e., if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}\dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)} = \beta_{\text{opt}}$

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



Summary

ML: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq 1$

MAP: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}$

MEG: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}\dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}$

The likelihood ratio is a "sufficient statistic".

Standardized SDT

None of the derivations so far made any assumptions about the signal and noise distributions (even though the graphs looked Gaussian). Thus, all statements I've made about ML/MAP/MEG are true for *any* distributions: discrete (such as Poisson) vs. continuous, unequal signal vs. noise distributions, univariate vs. multivariate. The likelihood principle still holds.

However, the standard SDT model that is most often used assumes equal-variance Gaussians:

$$p(x\,|\,N) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_N)^2}{2\sigma^2}\right) \text{ and } p(x\,|\,S) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_S)^2}{2\sigma^2}\right)$$

Standardized SDT

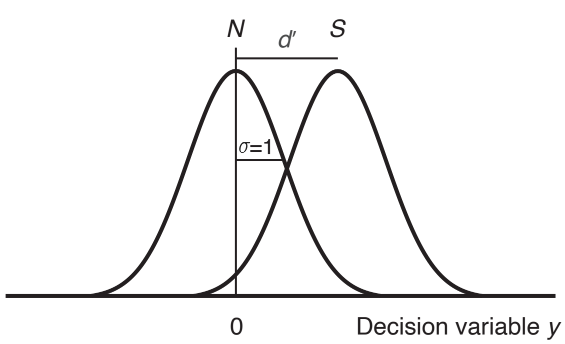
$$p(x\,|\,N) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_N)^2}{2\sigma^2}\right) \text{ and } p(x\,|\,S) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_S)^2}{2\sigma^2}\right)$$

Change of variables: Let $y = \dfrac{x-\mu_N}{\sigma}$, $d' = \dfrac{\mu_S-\mu_N}{\sigma}$

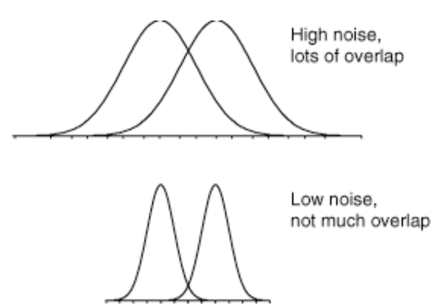
Then:

$$p(y\,|\,N) = \frac{1}{\sqrt{2\pi}}\exp\frac{-y^2}{2} \text{ and } p(y\,|\,S) = \frac{1}{\sqrt{2\pi}}\exp\frac{-(y-d')^2}{2}$$

Standardized SDT



Signal Detection Theory: discriminability (d')

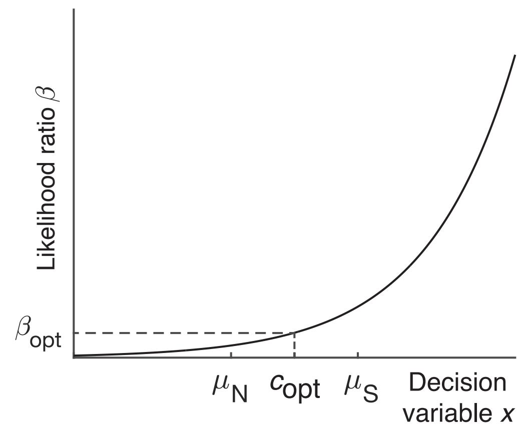


Standardized SDT

Likelihood ratio for $y = c$ is:

$$\frac{p(c\,|\,S)}{p(c\,|\,N)} = \frac{\exp\left[-\frac{(c-d')^2}{2}\right]}{\exp\left[-\frac{c^2}{2}\right]} = \exp\left[cd' - \frac{d'^2}{2}\right]$$

Standardized SDT



Standardized SDT

Likelihood ratio for $y = c$ is:

$$\beta_{\mathrm{opt}} = \frac{p(c_{\mathrm{opt}}\,|\,S)}{p(c_{\mathrm{opt}}\,|\,N)} = \frac{\exp\left[-\frac{(c_{\mathrm{opt}}-d')^2}{2}\right]}{\exp\left[-\frac{c_{\mathrm{opt}}^2}{2}\right]} = \exp\left[c_{\mathrm{opt}}d' - \frac{d'^2}{2}\right]$$

$$\log \beta_{\mathrm{opt}} = c_{\mathrm{opt}}d' - \frac{d'^2}{2}$$

$$c_{\mathrm{opt}} = \frac{d'}{2} + \frac{\log \beta_{\mathrm{opt}}}{d'}$$

$$= \frac{d'}{2} + \frac{1}{d'}\left[\log \frac{P(N)}{P(S)} + \log \frac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}\right]$$

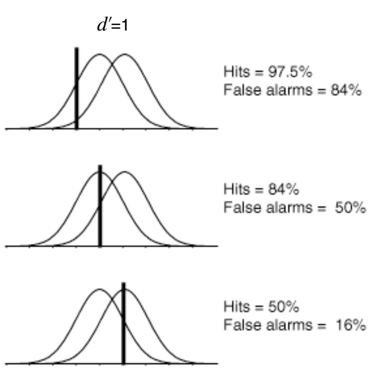
Standardized SDT

$$c_{\mathrm{opt}} = \frac{d'}{2} + \frac{1}{d'}\left[\log \frac{P(N)}{P(S)} + \log \frac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}\right]$$

and thus the optimal criterion is the ML criterion shifted by a term that is a function of the prior odds plus a term that is a function of the payoff ratio. Note that this additivity of the effects of priors and payoffs is *not* seen in human behavior.

Locke, S. M., Gaffin-Cahn, E., Hosseiniaveh, N., Mamassian, P. & Landy, M. S. (2020). *Attention, Perception, & Psychophysics*, *82*, 3158-3175.

Signal Detection Theory: Criterion



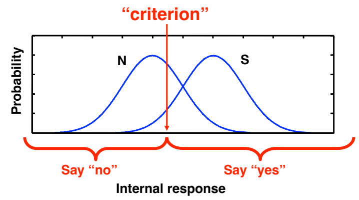
Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

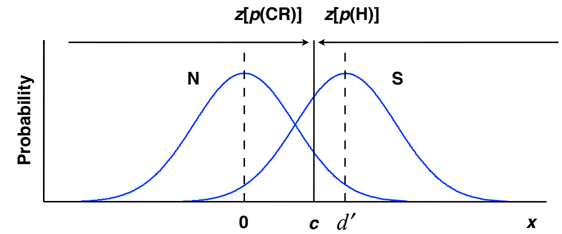
Example applications of SDT

- 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 Gaussian distributions, can we determine the underlying values of d' and criterion?



SDT: Estimating d' and c



$$
\begin{aligned}
d' &= z[P(\text{Hit})] + z[P(\text{Correct Reject})] \\
&= z[P(\text{Hit})] - z[P(\text{False Alarm})] \\
c &= z[P(\text{Correct Reject})], \text{ where}
\end{aligned}
$$

$$z(P) = \Phi^{-1}(P), \text{ where } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp \frac{z^2}{2} dz$$

SDT: Estimating d' and c

$$
\begin{aligned}
d' &= z[P(\text{Hit})] + z[P(\text{Correct Reject})] \\
&= z[P(\text{Hit})] - z[P(\text{False Alarm})] \\
c &= z[P(\text{Correct Reject})]
\end{aligned}
$$

Response

Stimulus	No	Yes
S	n_{Miss}	n_{Hit}
N	n_{CR}	n_{FA}

$$\hat{P}_{\text{Hit}} = n_{\text{Hit}} / (n_{\text{Hit}} + n_{\text{Miss}})$$
$$\hat{P}_{\text{FA}} = n_{\text{FA}} / (n_{\text{FA}} + n_{\text{CR}})$$

SDT: Estimating d' and c

$$
\begin{aligned}
d' &= z[P(\text{Hit})] + z[P(\text{Correct Reject})] \\
&= z[P(\text{Hit})] - z[P(\text{False Alarm})] \\
c &= z[P(\text{Correct Reject})]
\end{aligned}
$$

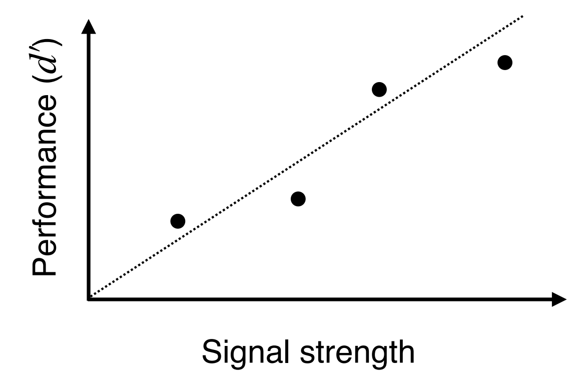
Response

Stimulus	No	Yes
S	$n_{\text{Miss}} + 0.5$	$n_{\text{Hit}} + 0.5$
N	$n_{\text{CR}} + 0.5$	$n_{\text{FA}} + 0.5$

$$\hat{P}_{\text{Hit}} = (n_{\text{Hit}} + 0.5)/(n_{\text{Hit}} + n_{\text{Miss}} + 1)$$
$$\hat{P}_{\text{FA}} = (n_{\text{FA}} + 0.5)/(n_{\text{FA}} + n_{\text{CR}} + 1)$$

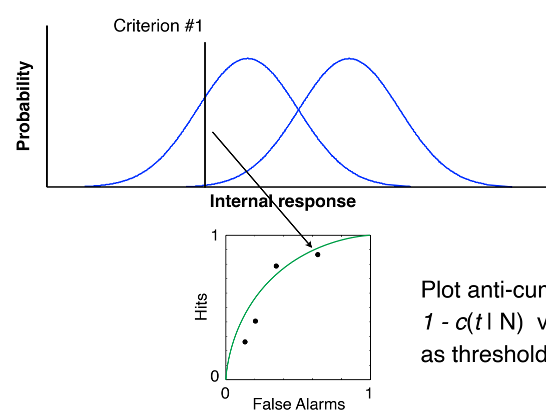
Hautus, M. J. (1995). Corrections for extreme proportions and the biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, 27, 46-51.

SDT: Psychometric function



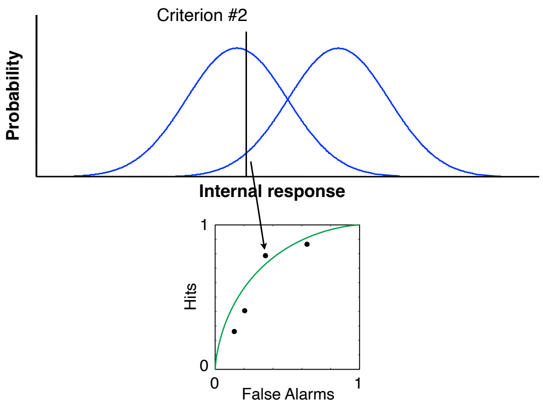
Note: 4 hit rates and one, shared, false-alarm rate

ROC (Receiver Operating Characteristic)

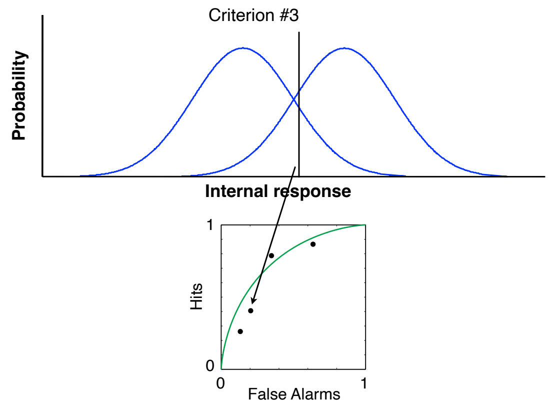


Plot anti-cumulatives:
1 - *c*(*t* | N) vs. *1*-*c*(*t* | *S*)
as threshold *t* varies

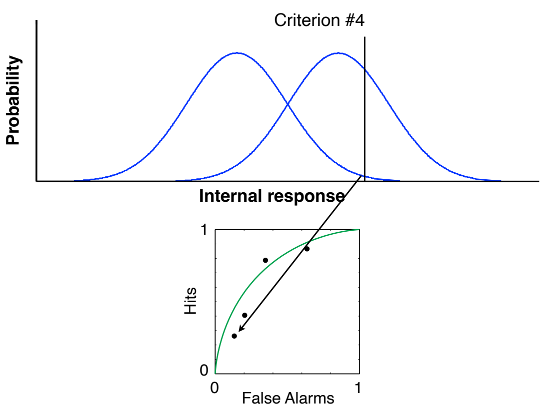
ROC (Receiver Operating Characteristic)



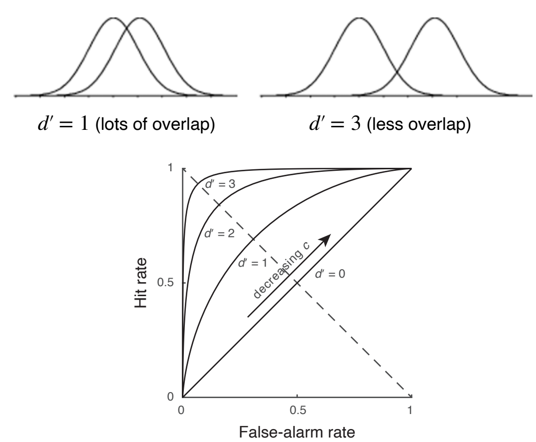
ROC (Receiver Operating Characteristic)



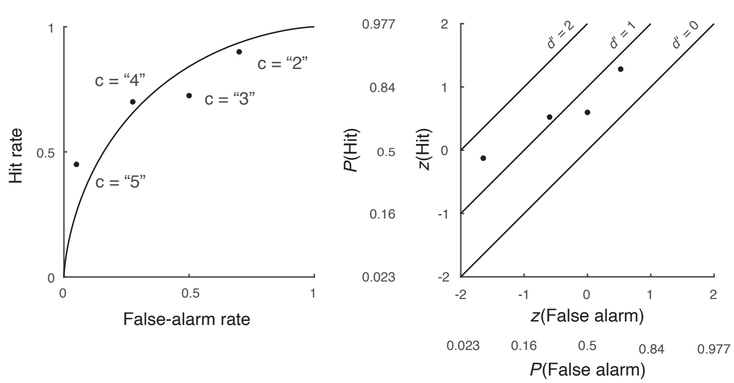
ROC (Receiver Operating Characteristic)



ROC (Receiver Operating Characteristic)



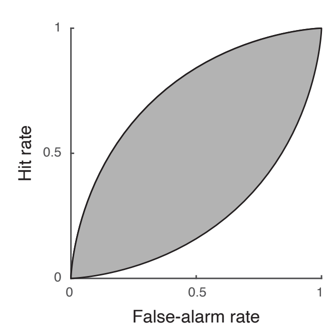
ROC (Receiver Operating Characteristic)



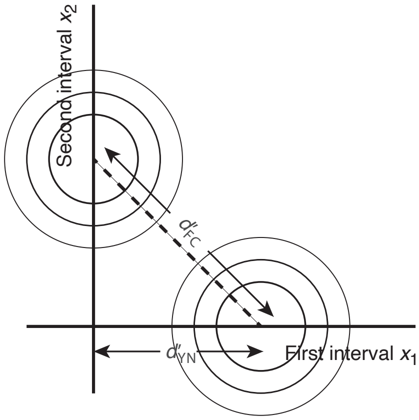
Or do a ML fit: Dorfman, D. D., & Alf, J., E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals: rating-method data. *Journal of Mathematical Psychology*, *6*, 487-496.

ROC (Receiver Operating Characteristic)

Achievable performance



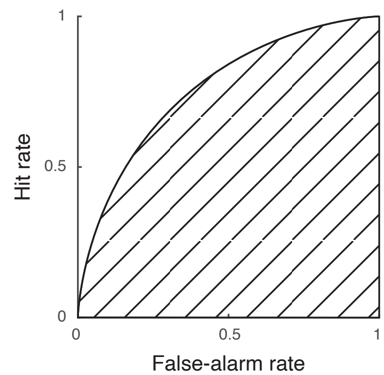
SDT and 2AFC



Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *48*, 1837-1851.

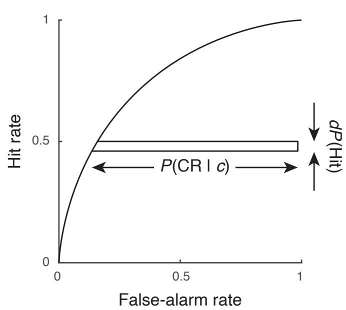
Area under the ROC

Area under curve = %correct in a 2AFC task



Area under the ROC

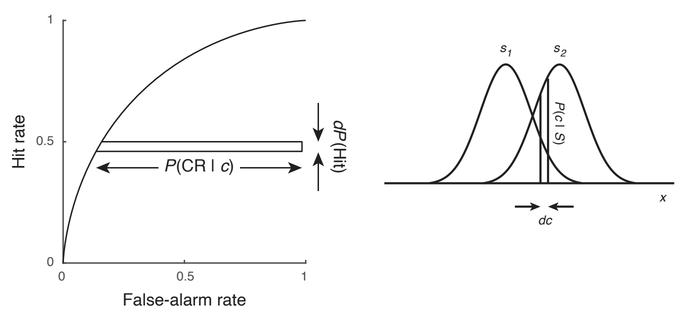
Area under curve = %correct in a 2AFC task



$$\text{AUROC} = \int_0^1 P(\text{Correct reject} \,|\, \text{criterion } c)\, dP(\text{Hit} \,|\, \text{criterion } c)$$

Area under the ROC

Area under curve = %correct in a 2AFC task



Slope of the ROC = likelihood ratio!

Area under the ROC

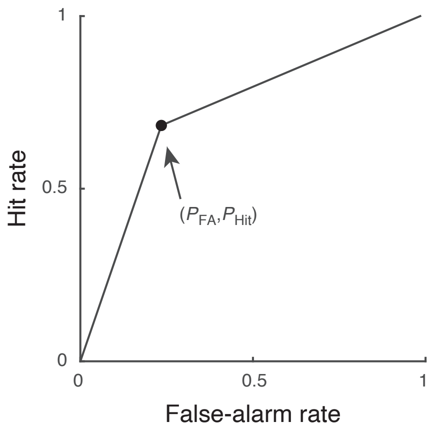
Area under curve = %correct in a 2AFC task

$$\begin{aligned}\text{AUROC} &= \int_0^1 P(\text{Correct reject} \,|\, \text{criterion } c)\, dP(\text{Hit} \,|\, \text{criterion } c)\\ &= \int_{-\infty}^{\infty} p(x < c \,|\, N)\, p(\text{measurement is } c \,|\, S)\, dc\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{c} p(x \,|\, N)\, p(\text{measurement is } c \,|\, S)\, dx\, dc\\ &= \int_{-\infty}^{\infty} p(\text{measurement is } c \,|\, S)\int_{-\infty}^{c} p(x \,|\, N)\, dx\, dc\\ &= P_{\text{FC}}\,.\end{aligned}$$

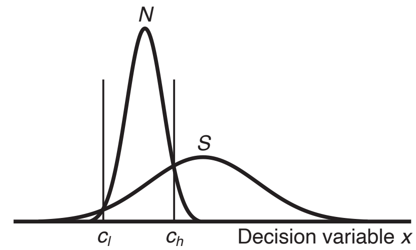
Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

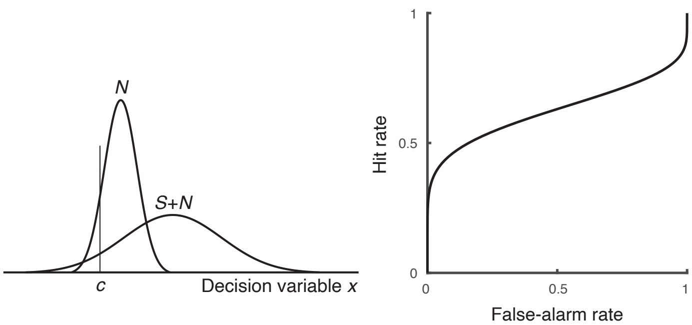
High-threshold theories



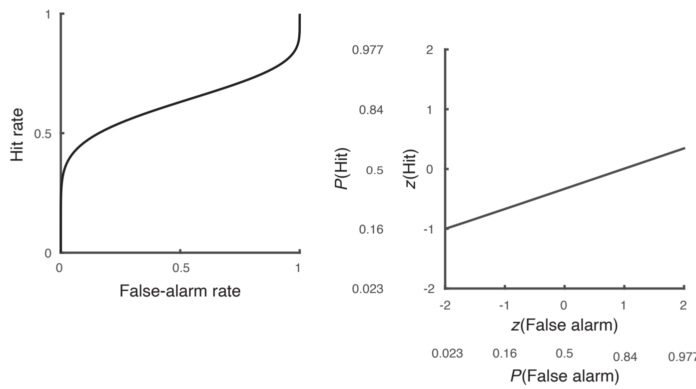
SDT with unequal variances



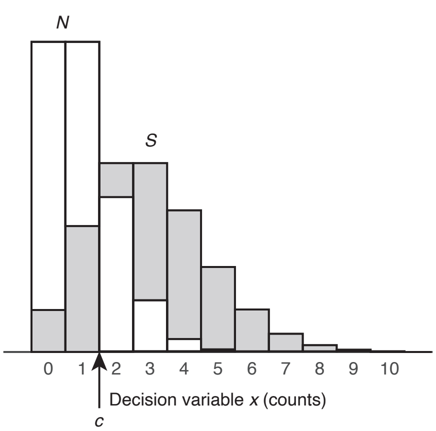
SDT with unequal variances



SDT with unequal variances



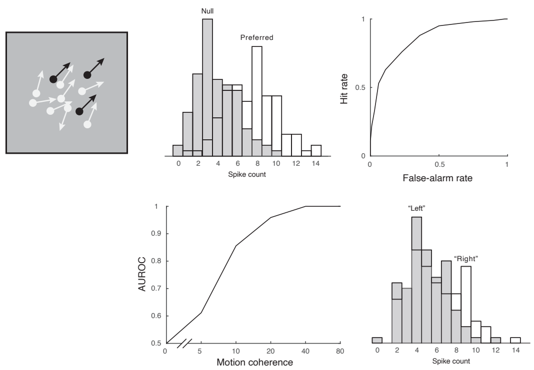
SDT with a discrete (Poisson) distribution



Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

Area under the ROC - Poisson case or with data: Neurometric function and choice probability

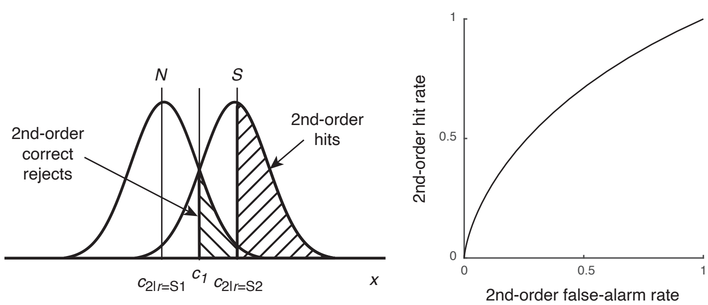


Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745-4765.
Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*, 87-100.

Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

Metacognition and SDT: Meta-d'



Metacognition and SDT: Meta-d'

Tasks:
- Confidence rating
- Estimate probability correct
- Confidence forced choice

Models:
- Meta-d'
- Casandre
- Confidence noise/confidence boost

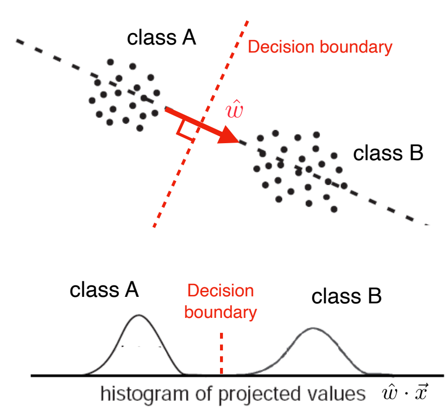
Rahnev, D. (2025). A comprehensive assessment of current methods for measuring metacognition. *Nature Communications*, *16*:701.
Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422-430.
Boundy-Singer, Z. M., Ziemba, C. M. & Goris, R. L. T. (2022). Confidence reflects a noisy decision reliability estimate. *Nature Human Behavior*, *7*, 142-154.
Mamassian, P., & de Gardelle, V. (2022). Modeling perceptual confidence and the confidence forced-choice paradigm. *Psychological Review*, *129*, 976–998.

Outline

1. Observer's perspective
2. Experimenter's perspective
3. Alternative theories or parameterizations
4. Applications to neural data
5. Applications to metacognition
6. Optional bonus: Multidimensional classifiers

Linear Classifier

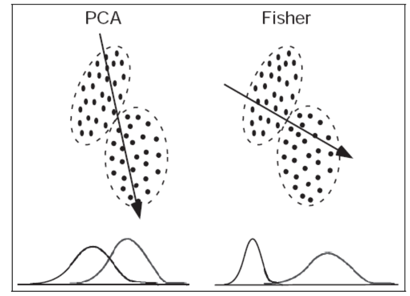
Find unit vector $\hat{w}$ ("discriminant") that best separates the distributions



Simplest linear discriminant: the Prototype Classifier

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

Fisher Linear Discriminant

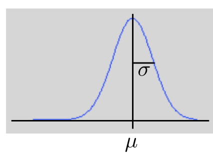


$$\max_{\hat{w}} \frac{\left[\hat{w}^T(\vec{u}_A - \vec{u}_B)\right]^2}{\left[\hat{w}^T C_A \hat{w} + \hat{w}^T C_B \hat{w}\right]}$$ (note: this is d' squared!)

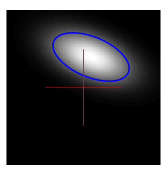
optimum: $\hat{w} = C^{-1}(\vec{u}_A - \vec{u}_B)$, where $C = \frac{1}{2}(C_A + C_B)$

Multi-dimensional Gaussian density

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



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

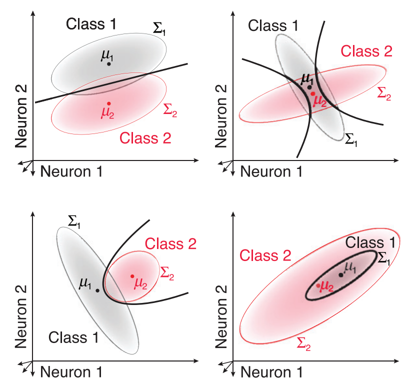


mean: [0.2, 0.8]
cov: [1.0 -0.3;
-0.3 0.4]

ML (or MAP) classifier for two Gaussians

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

Simplest case: equal covariances



[figure: Pagan et al. 2016]

Slides

Paper