

# ***B6: Statistics of Signal Detection Models***

Kenneth Knoblauch

Inserm U 846

Stem Cell and Brain Research Institute

Dept. Integrative Neurosciences

Bron, France

([ken.knoblauch@inserm.fr](mailto:ken.knoblauch@inserm.fr))



Bremen, August 25, 2013

# Outline

1. Structure of detection experiment
2. Equal Variance Gaussian SDT model
  - a. Closed form maximum likelihood solution
  - b. Direct Maximum likelihood estimates
  - c. Generalized Linear Model approach
3. Complete separation and the Hauck-Donner effect.
4. Forced-choice paradigms
5. Unequal Variance Gaussian SDT model
6. Rating scale paradigms and ROC curves
  - a. Ordinal regression with cumulative link models
7. Linear and Generalized Linear Mixed-effects models,
8. Maximum Likelihood Difference Scaling (if time)
9. Perspectives

**(seminal article)**

Tanner, W.P., Swets, J.A.: A decision-making theory of visual detection. *Psychological Review* **61**, 401-409 (1954)

**(seminal reference)**

Green, D.M., Swets, J.A.: *Signal Detection Theory and Psychophysics*. Robert E. Krieger Publishing Company, Huntington (1966/1974)

**(didactic references)**

Macmillan, N.A., Creelman, C.D.: *Detection Theory: A User's Guide*, Second ed.. Lawrence Erlbaum Associates, New York (2005)

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

McNicol, D. *A Primer of Signal Detection Theory*. Allen and Unwin (1972)

**(computational references)**

Knoblauch, K., Maloney L.T., *Modeling Psychophysical Data in R*. Springer, New York (2012)

Kingdom, F.A.A., Prins, N: *Psychophysics: A Practical Introduction*. Academic Press, New York (2009)

# Signal Detection Theory (SDT): some definitions

**Detection theory**, or **signal detection theory**, is a means to quantify the ability to discern between information-bearing patterns (called [stimulus](#) in humans, [signal](#) in machines) and random patterns that distract from the information (called [noise](#), consisting of background stimuli and random activity of the detection machine and of the nervous system of the operator).

([http://en.wikipedia.org/wiki/Detection\\_theory](http://en.wikipedia.org/wiki/Detection_theory))

Signal detection theory (SDT),...is a framework of statistical methods used to model how observers classify sensory events.

(Knoblauch, K. & Maloney, L. T. (2012) *Modeling Psychophysical Data in R*, Ch. 3)

Signal Detection Theory (SDT) was introduced...to address the problem of observer bias. In a simple Yes-No task, a psychometric function summarizes the probability of a 'Yes' response. However, this probability confounds the observer's ability to detect the presence or absence of a signal with any preference the observer may have to respond 'Yes'. Even when the signal intensity is zero (no signal is presented), the observer may still respond 'Yes'....The ... response measures are combined to allow separate estimates of a measure of the observer's *sensitivity* to the signal and the observer's *bias* (tendency to say "Present" independent of the presence or absence of the signal).

Knoblauch, K. & Maloney, L. T. (2012) *Modeling Psychophysical Data in R*, Preface)

# Detection Experiment (SDT)

Yes/No (Present/Absent) Experiment:

The Stimulus:

A signal is chosen by the experimenter and on a set of  $N$  trials, is divided randomly into  $r$  Signal-Present and  $N-r$  Signal-Absent trials (often  $r = N-r$ ).

The Task:

On each trial, the observer must make a decision (*classify* the trial) and respond as to whether the signal was *Present* or *Absent*.

This leads to 4 possible outcomes:

Stimulus Present / Response Present: Hit

Stimulus Present / Response Absent: Miss

Stimulus Absent / Response Present: False Alarm

Stimulus Absent / Response Absent: Correct Rejection

	Signal	Not Signal
Yes	Hit (H)	False Alarm (FA)
No	Miss (M)	Correct Rejection (CR)

## Response Classification Table

	Signal	Not Signal
Yes	Hit (H)	False Alarm (FA)
No	Miss (M)	Correct Rejection (CR)

Over  $N$  trials, we can calculate the proportion of trials that fall in each of the four categories to estimate the probabilities of each category:

	Signal	Not Signal
Yes	$P_H = P(Y \mid Present)$	$P_{FA} = P(Y \mid Absent)$
No	$P_M = P(No \mid Present)$	$P_{CR} = P(N \mid Absent)$

or

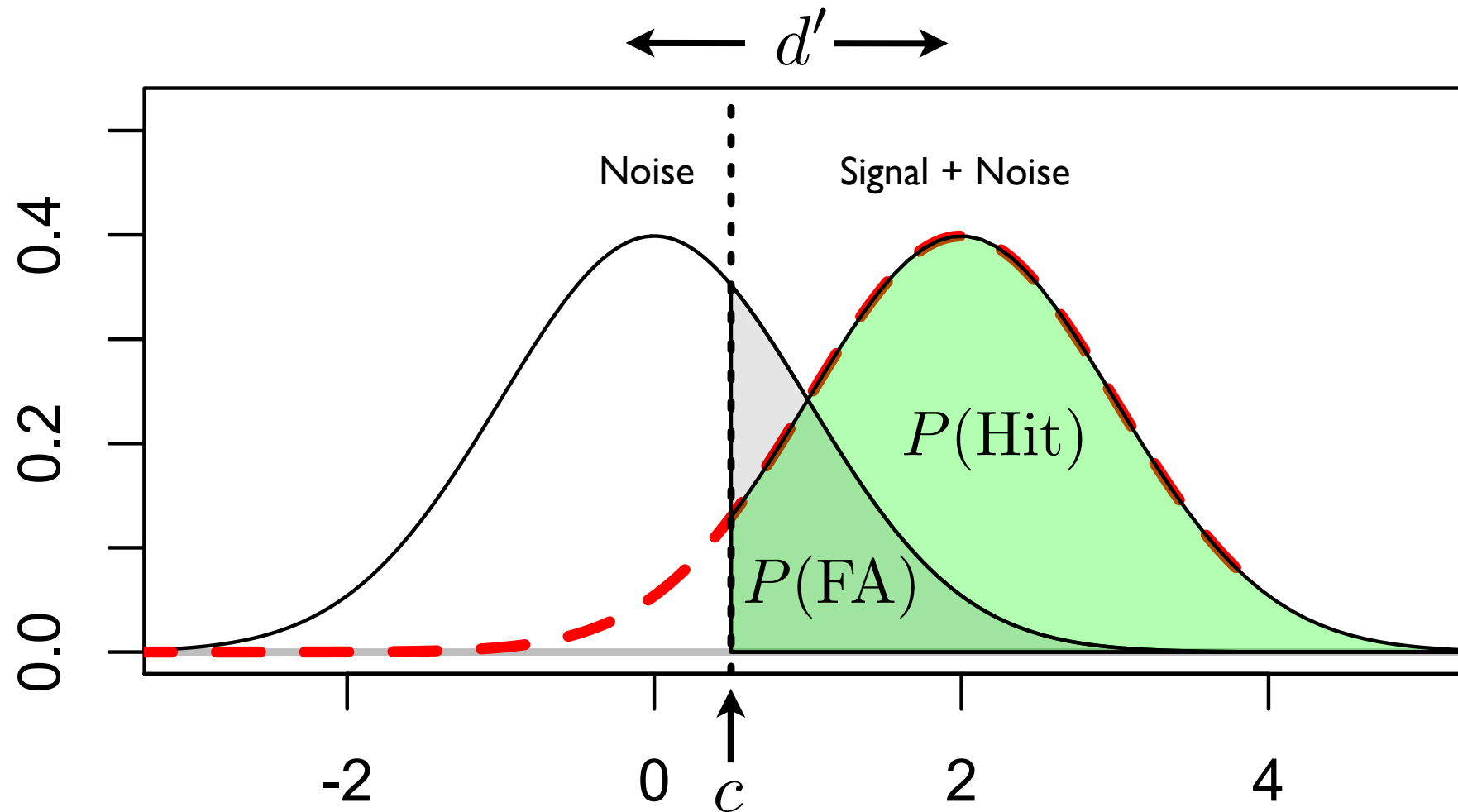
	Signal	Not Signal
Yes	$P_H = \#Yes / \#Present$	$P_{FA} = \#Yes / \#Absent$
No	$P_M = \#No / \#Present$	$P_{CR} = \#No / \#Absent$

Notez Bien:  $P_H = 1 - P_M$

$$P_{FA} = 1 - P_{CR}$$

# Signal Detection Theory: Equal Variance Gaussian Model

## Decision Space



## Internal Response

Decision Rule:

If response  $> c$ , choose “Yes”  
otherwise, choose “No”

Closed form (maximum likelihood) solution:

$$d' = \Phi^{-1}(P_H) - \Phi^{-1}(P_{FA})$$

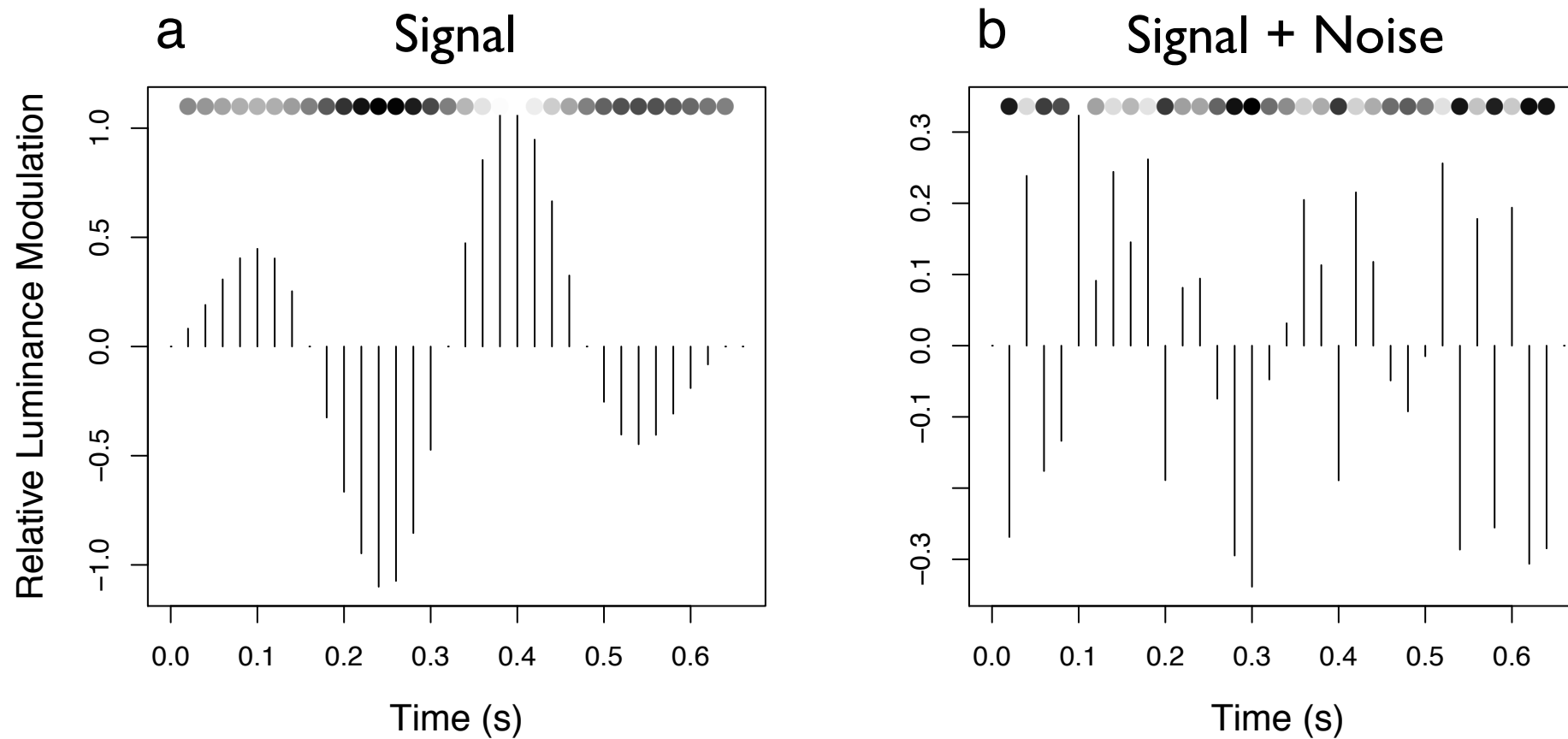
$$c = -0.5(\Phi^{-1}(P_H) + \Phi^{-1}(P_{FA}))$$

Go to shiny demo for ROC demo



# Example 1

Simple detection task: Gabor temporal modulation in noise



16 sessions of 224 trials (3584 total).

Signal present on half of the trials.

Hit =  $P(\text{Resp} = \text{Present} \mid \text{Stimulus} = \text{Present})$

Miss =  $P(\text{Resp} = \text{Absent} \mid \text{Stimulus} = \text{Present})$

False Alarm =  $P(\text{Resp} = \text{Present} \mid \text{Stimulus} = \text{Absent})$

Correct Rejection =  $P(\text{Resp} = \text{Absent} \mid \text{Stimulus} = \text{Absent})$

Stim\Resp	Absent	Present
Absent	CR	FA
Present	M	H

# Example 1

```
load("GabResp.Rdata") # load data file
str(GabResp)
Factor w/ 4 levels "H","FA","M","CR": 1 1 1 3 1 2 2 4 3 4 ...

table(GabResp) # cross-classify counts of each level
  H   FA   M   CR
1278 618 514 1174

# Restructure into data frame
Gb.df <- data.frame( Resp = GabResp %in% c("H", "FA"),
                     Stim = GabResp %in% c("H", "M"))

str(Gb.df)
'data.frame': 3584 obs. of 2 variables:
 $ Resp: logi TRUE TRUE TRUE FALSE TRUE TRUE ...
 $ Stim: logi TRUE TRUE TRUE TRUE TRUE FALSE ...

head(Gb.df) # TRUE = Present (1) ; FALSE = ABSENT (0)
  Resp Stim
1 TRUE TRUE
2 TRUE TRUE
3 TRUE TRUE
4 FALSE TRUE
5 TRUE TRUE
6 TRUE FALSE
...
```

# Example 1

```
( Gb.tab <- with(Gb.df, table(Stim, Resp)) ) # tabulate 4 cases
```

	Resp				
Stim	FALSE	TRUE	#	Absent	Present
FALSE	1174	618	#	CR	False Alarm
TRUE	514	1278	#	Miss	Hit

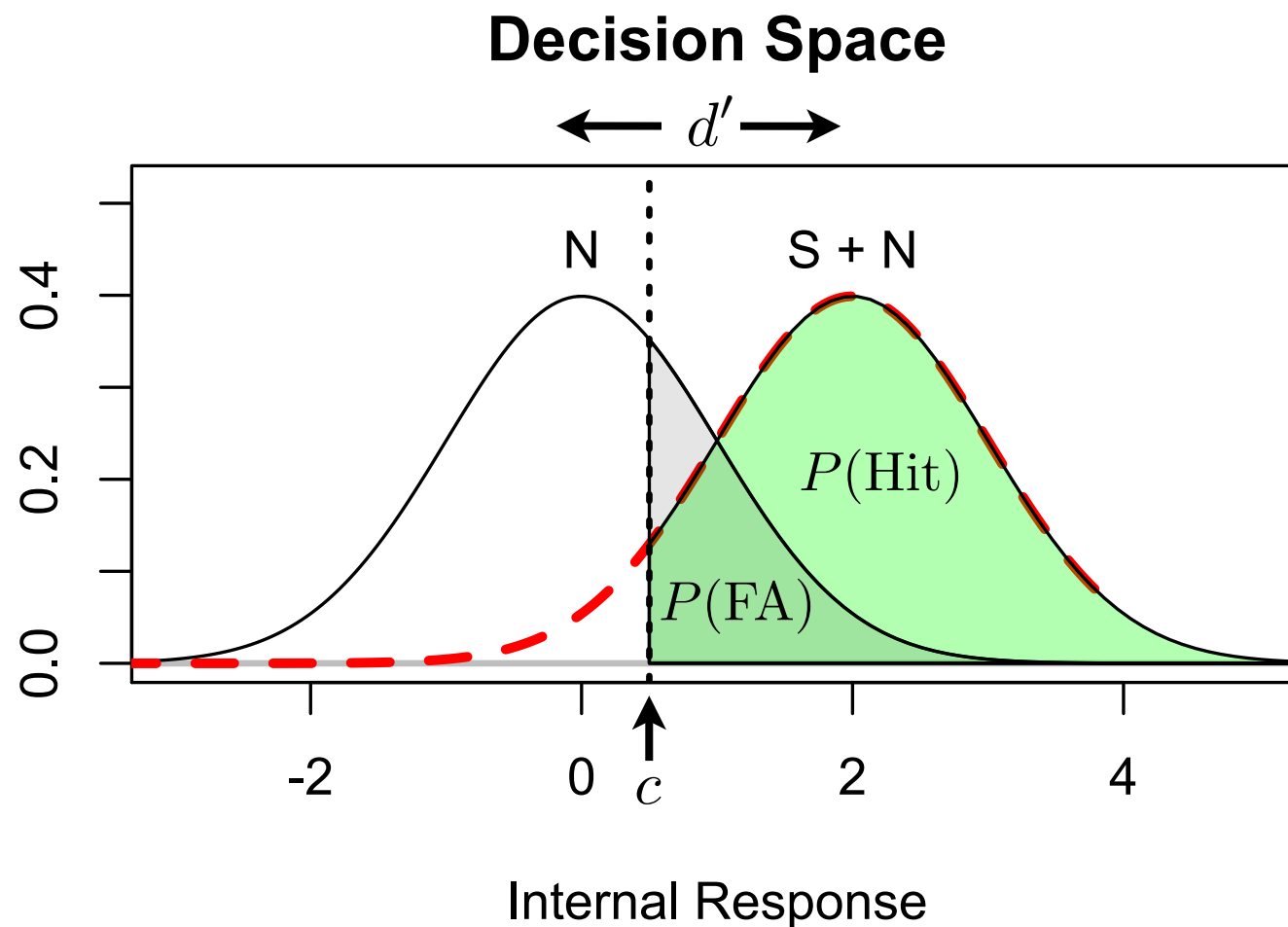
```
prop.table(Gb.tab, 1) # normalize on row sums
```

	Resp				
Stim	FALSE	TRUE	#	Absent	Present
FALSE	0.6551339	0.3448661	#	P(CR)	P(FA)
TRUE	0.2868304	0.7131696	#	P(M)	P(H)

```
diff(qnorm(prop.table(Gb.tab, 1)[, 2]))
```

```
TRUE  
0.961887 # d'
```

$$\Phi^{-1}(P_H) - \Phi^{-1}(P_{FA})$$



**Alternative parameterization:**

$$P_{FA} = 1 - \Phi(c)$$

$$c = \Phi^{-1}(1 - P_{FA}) = -\Phi^{-1}(P_{FA}) = \beta_0$$

$$d' = c - \Phi^{-1}(1 - P_H) = -\Phi^{-1}(P_{FA}) + \Phi^{-1}(P_H) = \beta_1$$

$$\text{Response} = \beta_0 + \beta_1 X,$$

where  $X = 1$  if response is “Present” and otherwise 0

# Matrix Representation

$$E(Y) = P(Y = 1) = \Phi(\mathbf{X}\beta)$$

$Y$  = vector of responses (0, 1, 1, 0, ...)

$$\beta = (\beta_0, \beta_1)$$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \end{pmatrix}$$

Design or model matrix with  
1 column for intercept and  
1 column for Absence/Presence of Stimulus

$$\Phi^{-1}(P(Y = 1)) = \mathbf{X}\beta = \beta_0 X^0 + \beta_1 X^1$$

# Maximum Likelihood Solution (explicit estimation)

Bernoulli Likelihood:  $p^y (1 - p)^{1-y}$

$$\mathcal{L}(\beta; Y, \mathbf{X}) = \prod_i \Phi(\beta_0 + \beta_1 X_i)^{Y_i} (1 - \Phi(\beta_0 + \beta_1 X_i))^{1-Y_i}$$

log likelihood to maximize over all experimental trials:

$$\begin{aligned} \ell(\beta; Y, \mathbf{X}) = \sum_i Y_i \log(\Phi(\beta_0 + \beta_1 X_i)) + \\ (1 - Y_i) \log(1 - \Phi(\beta_0 + \beta_1 X_i)) \end{aligned}$$

# Maximum Likelihood Solution

```
X <- model.matrix(~ Stim, data = Gb.df) # create model matrix
```

```
head(X)
```

	(Intercept)	StimTRUE
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	0

...

```
llik <- function(b, X, Y){ # function to compute log likelihood
```

```
  p <- pnorm(X %*% b)
```

```
  -sum(Y * log(p) + (1 - Y) * log(1 - p))
```

```
}
```

```
optim(par = c(0, 1), fn = llik, X = X, Y = Gb.df$Resp) # minimize -log Likelihood
```

```
$par
```

```
[1] -0.3991859  0.9617757      # estimated parameters
```

```
$value
```

```
[1] 2228.351      # -log Likelihood
```

...

# Maximum Likelihood Solution

```
optim(c(0, 1), llik, llik, X = X, Y = Gb.df$Resp)$par  
[1] -0.3991859  0.9617757  # d' differs a little from closed form estimate  
  
# insist on stricter relative tolerance at solution  
optim(c(0, 1), llik, X = X, Y = Gb.df$Resp, control = list(reltol = 1e-12))$par  
[1] -0.3992183  0.9618875  # This now matches closed form solution  
  
# Hessian matrix is the matrix of second partial derivatives of log Likelihood  
  
Gb.opt <- optim(c(0, 1), llik, X = X, Y = Gb.df$Resp, control = list(reltol = 1e-12),  
               hessian = TRUE) # add argument to calculate Hessian matrix at ML solution  
  
Gb.opt$par # estimated parameters  
[1] -0.3992183  0.9618875  
  
# approximate standard errors obtained from  
# square root of diagonal elements of inverse of Hessian matrix  
sqrt(diag(solve(Gb.opt$hessian)))  
0.03048029 0.04374257
```



# The *Generalized* Linear Model (GLM) Solution

(not general linear model!!!!)

1. Response is distributed as an exponential family member  
(Binomial, Poisson, Gamma, Gaussian, ...)?

$$f(y; \theta, \phi) = \exp \left[ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \theta) \right]$$

2. Linear predictor, specifying how the explanatory variables enter the model

$$\eta = X\beta$$

3. Link function,  $g$ , relating the mean to the linear predictor

$$g(\mu) = g[E(Y)] = \eta = X\beta$$

McCullagh P. and Nelder, J. A. (1989) *Generalized Linear Models*. London: Chapman and Hall.

Dobson, A. J. (1990) *An Introduction to Generalized Linear Models*. London: Chapman and Hall.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, **3**, 186-205.

Knoblauch, K., Maloney L.T., *Modeling Psychophysical Data in R*. Springer, New York (2012)

# Exponential family

The Gaussian, Poisson and Binomial (and others) are members of the exponential family.

$$f(y; \theta, \phi) = \exp \left[ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \theta) \right]$$

Gaussian:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{\sigma^2} \left( -x_i\mu + \frac{\mu^2}{2} + \frac{x_i^2}{2} \right) \right]$$

canonical link is identity function.

Binomial:

$$f(y; p, n) = \binom{n}{r} p^r (1 - p)^{n-r} = \binom{n}{r} \exp \left[ r \log \left( \frac{p}{1 - p} \right) + n \log(1 - p) \right]$$

canonical link is log odds ratio (logit) function, but for EVG model we use probit,  $\Phi^{-1}$ .

Simple expressions for the likelihood, mean and variance that can be just plugged in and then the model fit by standard procedures.

# GLM solution

```
head(Gb.df)
```

```
  Resp Stim
1  TRUE TRUE
2  TRUE TRUE
3  TRUE TRUE
4 FALSE TRUE
5  TRUE TRUE
6  TRUE FALSE
```

Resp is a binary response variable

Stim could be considered as a 2-level factor.

Formula Object:  $\text{Resp} \sim \text{Stim}$

Linear Predictor:

$$\text{Resp} = \beta_0 + \beta_1 \text{Stim}$$

Stim0:  $\beta_0 = \Phi^{-1}(P_{\text{FA}})$

Stim1 - Stim0:  $\beta_1 = \Phi^{-1}(P_{\text{H}}) - \Phi^{-1}(P_{\text{FA}}) = d'$

```
Gb.glm <- glm(Resp ~ Stim, binomial(probit), Gb.df) #specify binomial family and probit link
summary(Gb.glm)
```

Call:

```
glm(formula = Resp ~ Stim, family = binomial(probit), data = Gb.df)
```

.....

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.39922	0.03048	-13.10	<2e-16	***
StimTRUE	0.96189	0.04374	21.99	<2e-16	*** # d', compare w/ previous calculations

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4956.4 on 3583 degrees of freedom  
Residual deviance: 4456.7 on 3582 degrees of freedom # twice the log Lik obtained with optim  
AIC: 4460.7 # Deviance  $(-2 \log \text{Lik}) + 2 * \text{\#estimated parameters (here, 2)}$

# Exercises

1. Gb.df is based on 16 sessions of 224 trials each. Estimate sensitivity and bias for each session individually and plot them as a function of session. Are there any learning trends? How variable are the measures?
2. Fit the SDT detection model to the Gb.df data set cumulatively by session, i.e., first the first session, then the first and second sessions, etc.  
How do the SE's of the parameter estimates vary with the number of trials?

# GLM solution: More complex experimental design

Observers shown 100 images in each of two sessions.

In second session, 50 images were New and 50 were Old

Two groups: Normal Observer, Hypnotized Observers

Example from Macmillan, N.A., Creelman, C.D. (2005)

Does hypnotism influence recognition memory?

```
Hypno.df <- read.table(textConnection("
```

	Stim	Number	Resp	Cond
1	Old	69	Yes	Normal
2	New	31	Yes	Normal
3	Old	31	No	Normal
4	New	69	No	Normal
5	Old	89	Yes	Hypnotized
6	New	59	Yes	Hypnotized
7	Old	11	No	Hypnotized
8	New	41	No	Hypnotized

Here we treat aggregated binomial responses instead of individual Bernoulli responses.

```
), TRUE)
```

```
Hypno.df$Cond <- relevel(Hypno.df$Cond, "Normal")
```

```
str(Hypno.df)
```

```
'data.frame': 8 obs. of 4 variables:
```

```
$ Stim : Factor w/ 2 levels "New","Old": 2 1 2 1 2 1 2 1
```

```
$ Number: int 69 31 31 69 89 59 11 41
```

```
$ Resp : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 2 1 1
```

```
$ Cond : Factor w/ 2 levels "Normal","Hypnotized": 1 1 1 1 2 2 2 2
```

```

m2 <- glm(Resp ~ Stim * Cond, binomial(probit), Hypno.df,
  weights = Number) # ~ Stim + Cond + Stim:Cond,
m1 <- update(m2, . ~ . - Stim:Cond) # remove interaction of Stim and Cond
m0 <- update(m1, . ~ . - Cond) # remove main effect of Cond
anova(m0, m1, m2, test = "Chisq") # nested likelihood ratio tests

```

Analysis of Deviance Table

$$-2(\ell_0 - \ell_1) \sim \chi^2_{df_0 - df_1}$$

Model 1: Resp ~ Stim

Model 2: Resp ~ Stim + Cond

Model 3: Resp ~ Stim \* Cond

	Resid.	Df	Resid.	Dev	Df	Deviance	Pr(>Chi)
1	6		480.84				
2	5		452.32	1	28.5225	9.262e-08	***
3	4		452.32	1	0.0007	0.9792	

```
library(effects)
```

```
allEffects(m2)
```

```
model: Resp ~ Stim * Cond
```

Stim\*Cond effect

Cond

Stim	Hypnotized	Normal
New	0.31	0.59
Old	0.69	0.89

```
summary(m2)
```

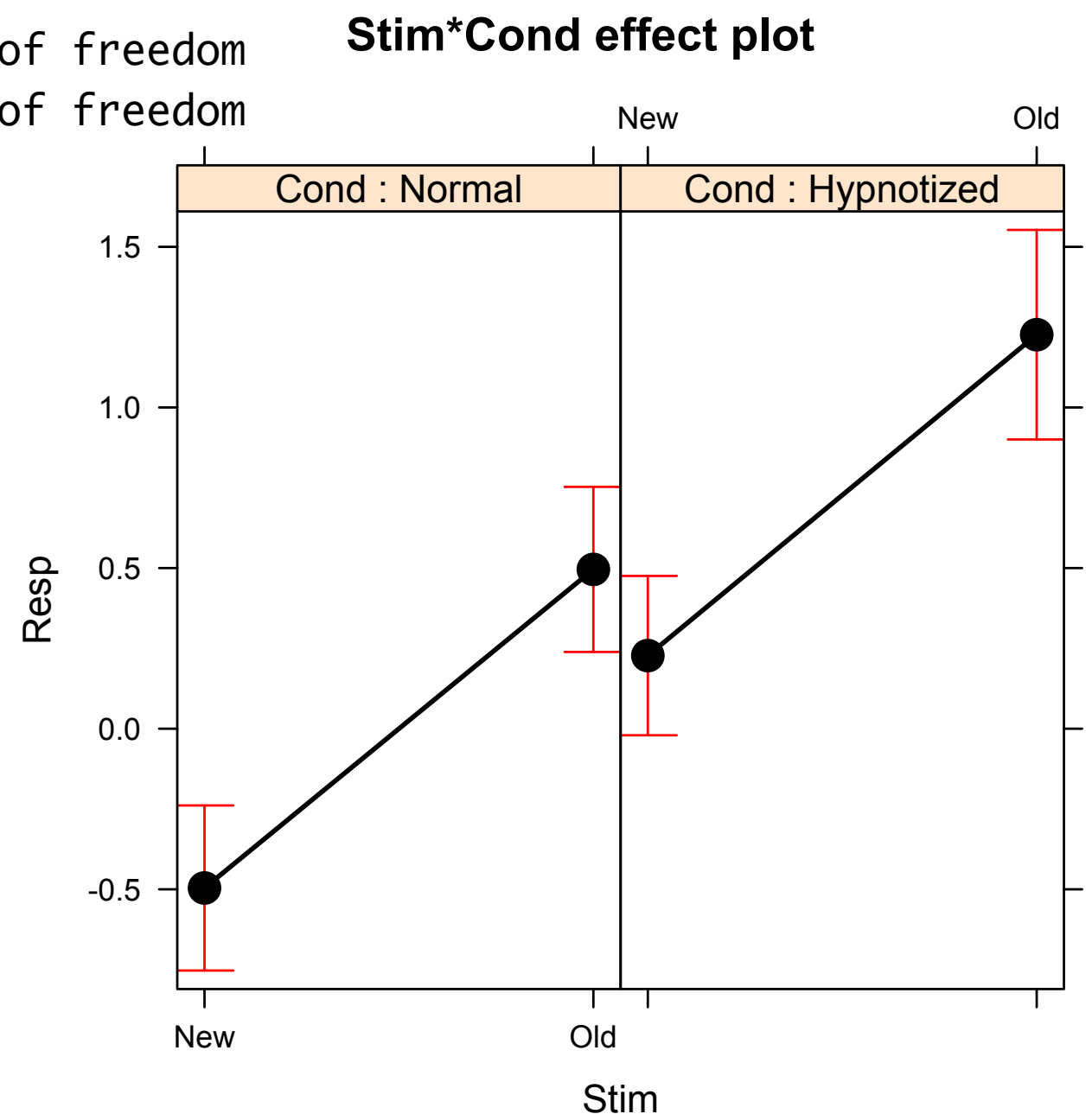
```
.....  
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.495850	0.131095	-3.782	0.000155	***
StimOld	0.991701	0.185396	5.349	8.84e-08	***
CondHypnotized	0.723395	0.182188	3.971	7.17e-05	***
StimOld:CondHypnotized	0.007282	0.279403	0.026	0.979206	

```
---
```

```
.....  
Null deviance: 531.25 on 7 degrees of freedom  
Residual deviance: 452.32 on 4 degrees of freedom  
AIC: 460.32
```

```
plot(allEffects(m2), type = "link")
```

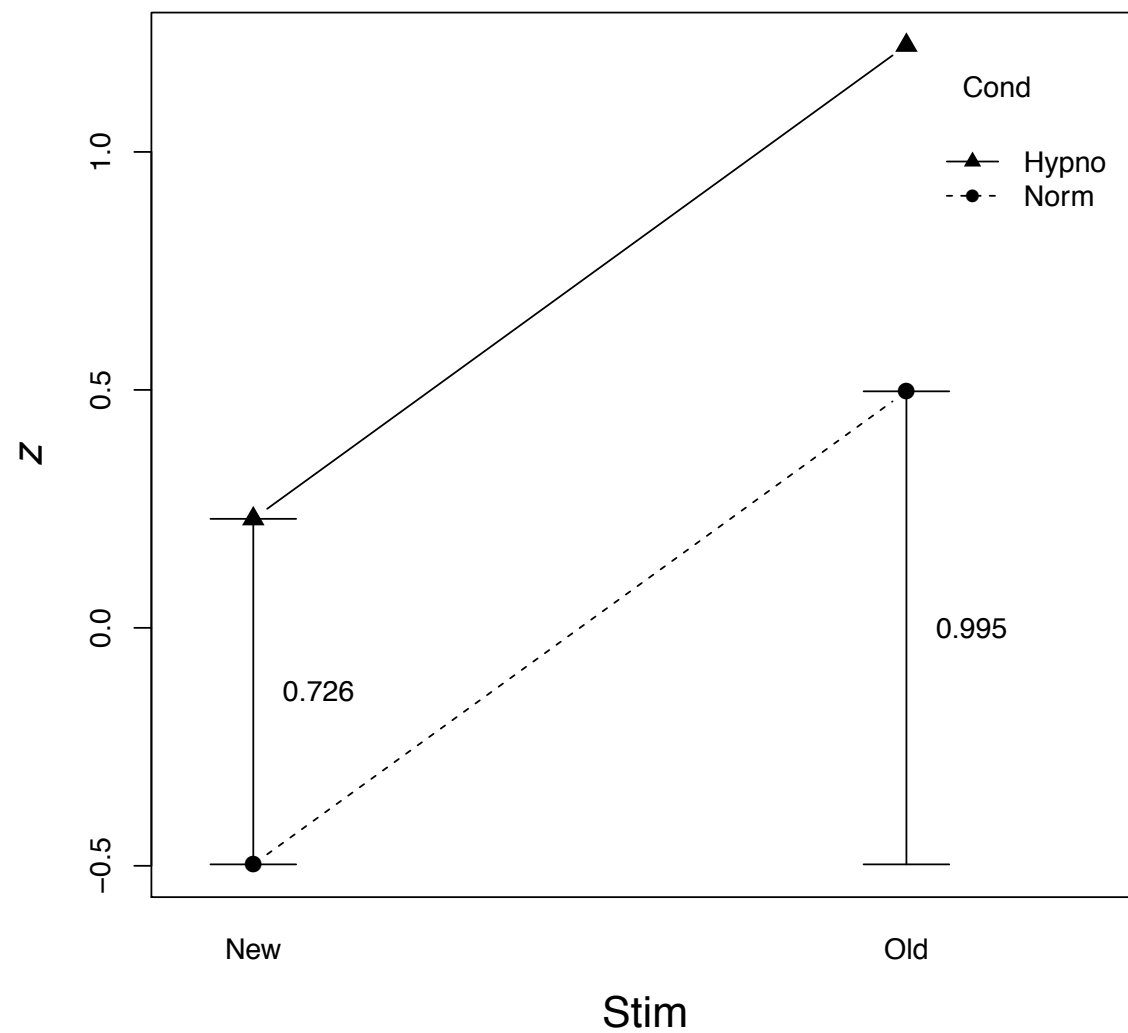
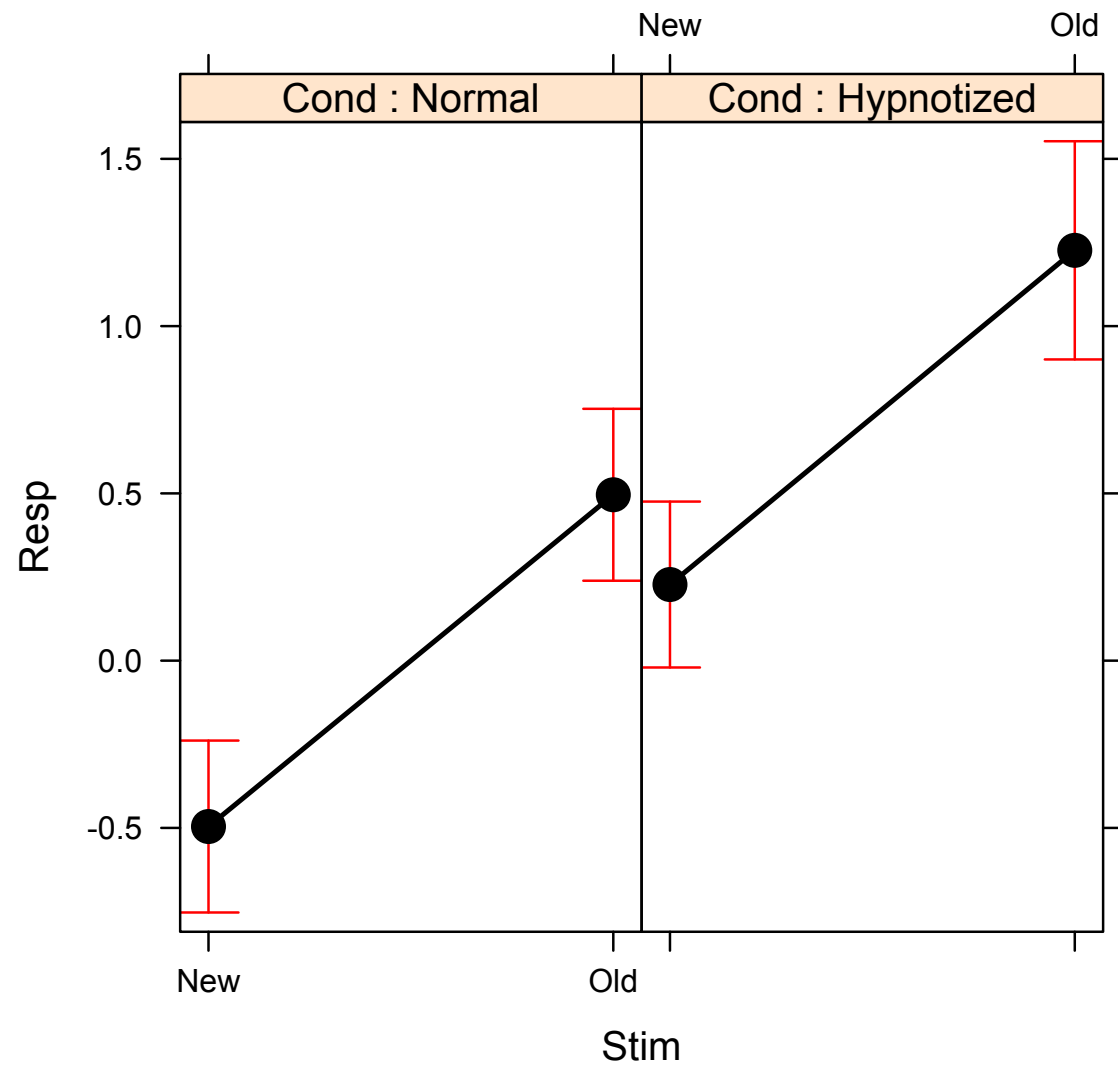


## Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.495850	0.131095	-3.782	0.000155	***
StimOld	0.991701	0.185396	5.349	8.84e-08	***
CondHypnotized	0.723395	0.182188	3.971	7.17e-05	***
StimOld:CondHypnotized	0.007282	0.279403	0.026	0.979206	

---

## Stim\*Cond effect plot





The preceding example was an ANOVA type design  
(all factors, i.e., categorical variables).

By using a covariate (continuous predictor variable, like contrast,  
distance, etc.),  
we can fit psychometric functions, i.e., a regression type design.

By mixing factors and covariates, we have an ANCOVA type design  
and can compare psychometric functions for different conditions,  
compare slopes of psychometric functions, etc.

# Complete separation and Hauck-Donner phenomenon

```
oops <- options(warn = 0) # re-setting to default for warnings
CS.df <- data.frame(Present = c(10000, 100),
                   Absent = c(0, 9900),
                   Stim = factor(c(1, 0)))
```

CS.df

	Present	Absent	Stim	
1	10000	0	1	
2	100	9900	0	# a bit lopsided but Oh, well...

```
summary(glm(cbind(Present, Absent) ~ Stim, binomial, CS.df))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-4.595e+00	1.005e-01	-45.72	<2e-16	***
Stim1	3.650e+01	6.711e+05	0.00	1	# d' estimate
	...				

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

**Complete separation** caused by 0 Absent for level Stim = 1,  
but note the estimate of  $d'$  ( $> 30$ ), and its SE ( $> 10^5$ ).

This is called the Hauck-Donner phenomenon, in which binomial models with large effects can have even larger standard errors, a frequent symptom of complete separation.

In fitting psychometric functions (having a covariate predictor) this arises with infinitely steep psychometric functions.

Hauck Jr, W.W., Donner, A.: Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association 72, 851–853 (1977)

# Complete separation and Hauck-Donner phenomenon

Changing the link might help. The preceding example used the default logit link, but if we switch to the probit link (that we have been using for its relation to EVG SDT) we obtain no warning but still large SE (try this on your own).

Just one Miss is sufficient to mitigate the problem:

```
CS1.df <- data.frame(Present = c(10000, 100),  
                     Absent = c(1, 9900),  
                     Stim = factor(c(1, 0)))
```

	Present	Absent	Stim
1	10000	1	1
2	100	9900	0

```
summary(glm(cbind(Present, Absent) ~ Stim, binomial, CS1.df))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-4.5951	0.1005	-45.72	<2e-16	***
Stim1	13.8055	1.0051	13.74	<2e-16	***

No warning and SE's have become more reasonable.

Moral: Avoid conditions of certain detection and collect enough data to avoid observing 0 cases.

# Forced-choice paradigms

When the observer the signal is 1 of  $m$  stimuli presented ( $m$  spatial positions,  $m$  intervals) we speak of  $m$ -alternative forced-choice (mAFC) or  $m$ -interval forced-choice (mIFC). It has been reputed that such procedures are less prone to bias<sup>1</sup>. Closed form solutions exist for 2AFC (2IFC), but not for other values of  $m$ . For an unbiased observer<sup>2</sup>, the probability of a correct response  $P_c$  is given by

$$P_c = \int_{-\infty}^{\infty} \phi(x - d') \Phi(x)^{m-1} dx$$

where  $\phi$  and  $\Phi$  are the normal density and distribution functions, respectively. Given the  $P_c$ ,  $d'$  may be calculated by finding the root of

$$f(d') = P_c - \int_{-\infty}^{\infty} \phi(x - d') \Phi(x)^{m-1} dx$$

<sup>1</sup>but see Yeshurun, Y., Carrasco, M., Maloney, L.T.: Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research* **48**, 1837–1851 (2008).

<sup>2</sup> for estimation for a biased observer see DeCarlo, L.T. (2012). On a signal detection approach to  $m$ -alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, 56, 196-207.

# Forced-choice paradigms

From **psyphy** package:

```
dprime.mAFC <- function (Pc, m)
{
  m <- as.integer(m)
  if (m < 2)
    stop("m must be an integer greater than 1")
  if (!is.integer(m))
    stop("m must be an integer")
  if (Pc <= 0 || Pc >= 1)
    stop("Pc must be in (0,1)")
  est.dp <- function(dp) {
    pr <- function(x) dnorm(x - dp) * pnorm(x)^(m - 1) # function to integrate
    Pc - integrate(pr, lower = -Inf, upper = Inf)$value # diff from target value
  }
  dp.res <- uniroot(est.dp, interval = c(-10, 10)) # find zero-crossing, ie, root
  dp.res$root
}
```

```
library(psyphy)
x <- c(0.25, 0.5, 0.75, 0.95)
sapply(x, dprime.mAFC, m = 4)
[1] -1.532277e-06  8.367746e-01  1.682178e+00  2.916233e+00
```

Go to shiny demo for unequal variance Gaussian demo

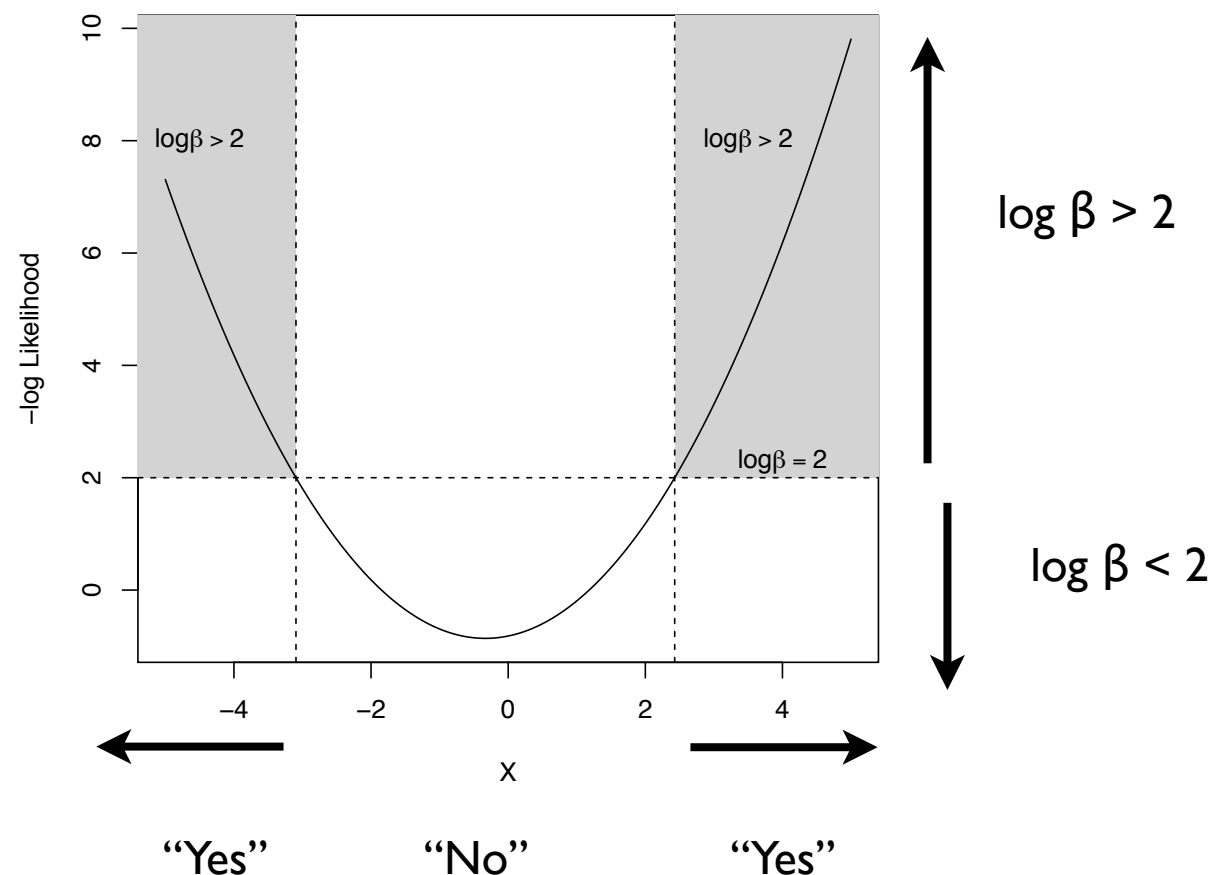
# Unequal Variance Gaussian Model

Likelihood ratio compared to criterion  $\beta$ :

$$\Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-d')^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)}$$

log Likelihood ratio is a quadratic function of  $x$ :

$$\lambda(x) = -\frac{1}{2\sigma^2} \left[ (1 - \sigma^2)x^2 - 2d'x + d'^2 + 2\sigma^2 \log(\sigma) \right] > \log(\beta)$$



Two domains of  $x$  (grey regions) that lead to a “Yes” response ( $\sigma > 1$ ).

In general, need to estimate ( $P_H$ ,  $P_{FA}$ ) for at least 2 criterion levels in order to solve the likelihood uniquely for  $d'$ .

Must solve likelihood directly by optimization. No closed form or glm solution.

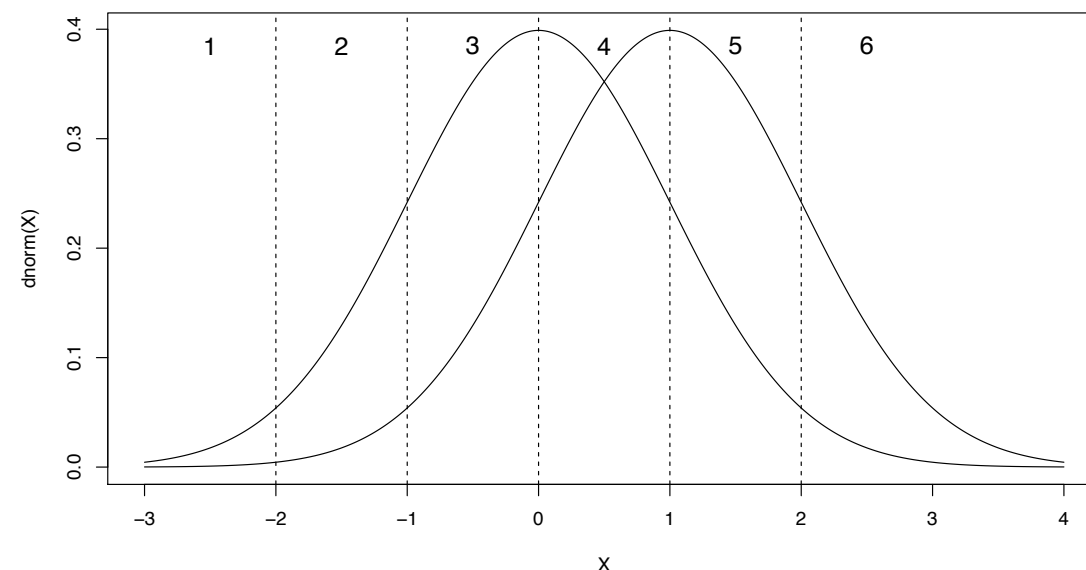
# Rating Scale Paradigms

To test equal variance assumption, we need to evaluate the Hit and False Alarm rates for 2 or more criterion levels. One could run several experiments, inducing the observer to change criteria across experiments (e.g., modify the reward/punishments for generating Hits and False Alarms, modifying the a priori probabilities of Signal Present and Absent).

Using a rating scale is a more efficient procedure that entails requiring the observer to use several criteria at once.

Run experiment as a simple detection task with Signal and Blank trials randomly interspersed. On each trial, ask the observer to rate his/her confidence on an n-point scale that the signal was present, e.g. for a 6 point scale,

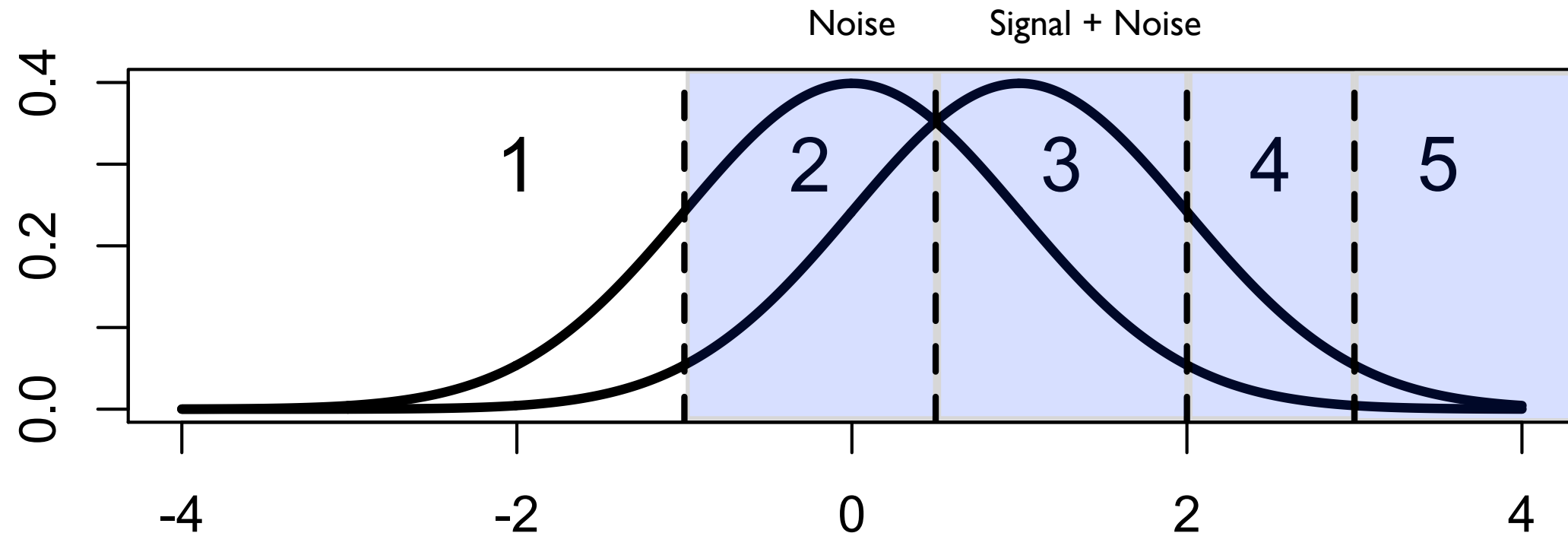
- 1, highly certain that signal absent
2. fairly certain that signal absent
3. not sure but probably absent
4. not sure but probably present
5. fairly certain that signal present
6. highly certain that signal present



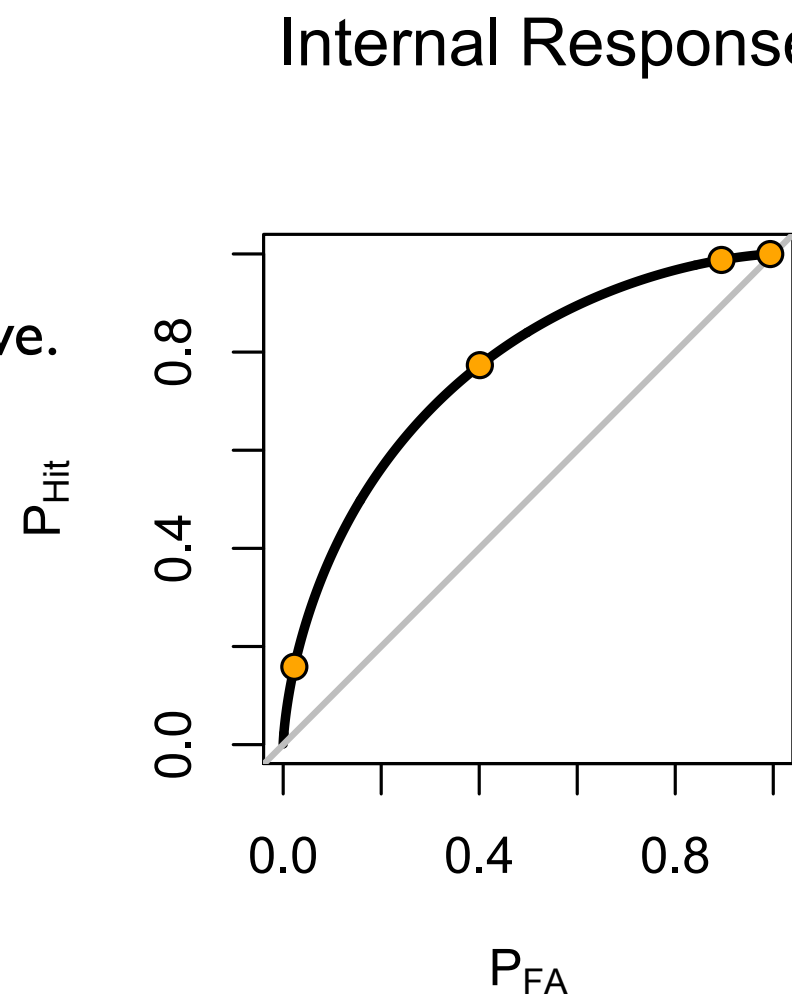
Hypothesis: Observer uses 6 criterion values to partition decision space.



# Analysis of rating scales: Signal Detection Theory



5 level rating scale:  
Hit and False alarm rate  
calculated for 4 of the 5  
criterion levels yielding  
4 points of the ROC curve.



ROC curve analysis

# Analysis of rating scales: ROC curve



How confident are you  
that these are siblings?

Maloney, L. T., and Dal Martello, M. F. (2006).  
Kin recognition and the perceived facial similarity  
of children. *Journal of Vision*, 6(10):4, 1047–1056.

11 point rating scale used: 0 definitely not, ..., 10 definitely

32 observers

30 images each

We'll ignore the observer differences and treat data set as one big observer.

# Analysis of rating scales: ROC curve

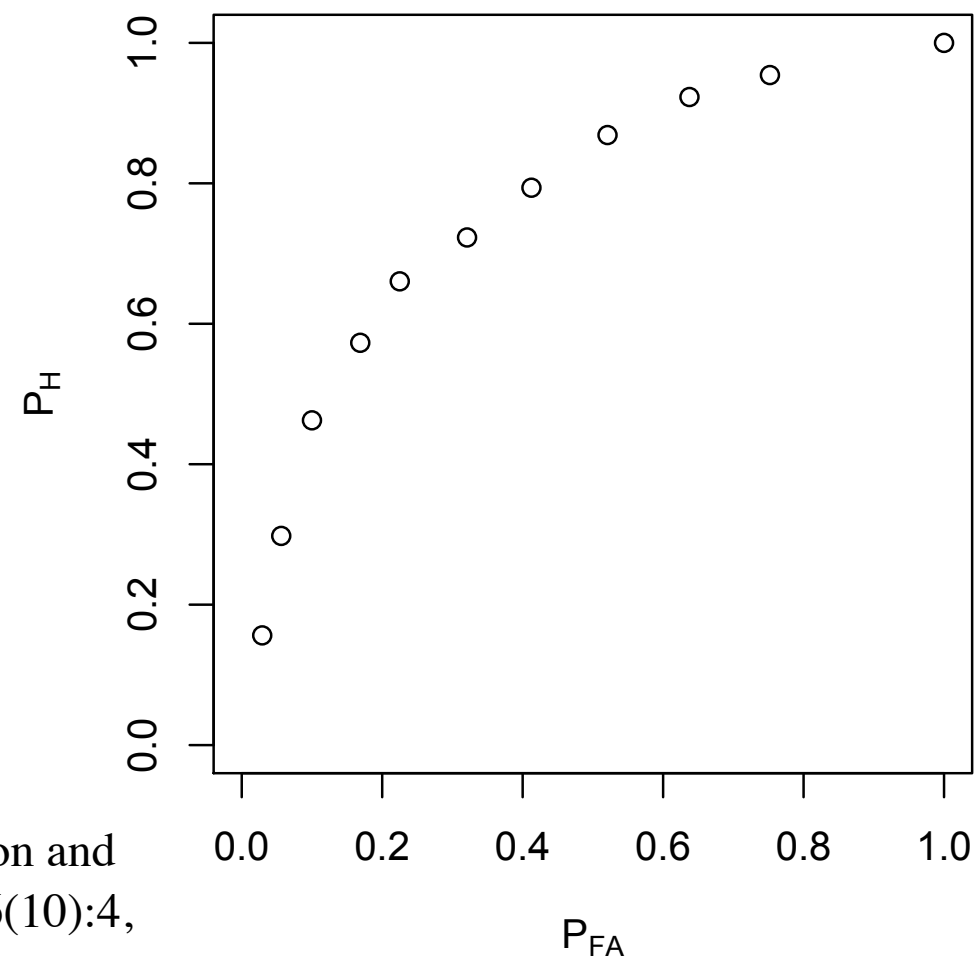
```
load("Faces.Rdata") # 11-point rating scale of whether a pair of faces
                    # belong to a pair of siblings or not
Faces.tab <- xtabs(~ sibs + SimRating, Faces) # tabulate sib status w/ ratings
```

```
      SimRating
sibs   0  1  2  3  4  5  6  7  8  9 10
  0 119 55 56 52 44 46 27 33 21 13 14
  1  22 15 26 36 34 30 42 53 79 68 75
```

```
Faces.csum <- apply(Faces.tab[, 11:1], 1, cumsum)[11:1, ]
Faces.ROC <- Faces.csum/Faces.csum[1, ] # cumulate and normalize
```

```
      sibs
      0      1
  0 1.00000000 1.00000000
  1 0.75208333 0.9541667
  2 0.63750000 0.9229167
  3 0.52083333 0.8687500
  4 0.41250000 0.7937500
  5 0.32083333 0.7229167
  6 0.22500000 0.6604167
  7 0.16875000 0.5729167
  8 0.10000000 0.4625000
  9 0.05625000 0.2979167
 10 0.02916667 0.1562500
```

ROC curve



Maloney, L. T., and Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision*, 6(10):4, 1047–1056.

# Analysis of rating scales: ROC curve

```
Faces.zROC <- qnorm(Faces.ROC)
```

```
sibs
```

	0	1
0	Inf	Inf
1	0.68106030	1.68667082
2	0.35178434	1.42496726
3	0.05224518	1.12050177
4	-0.22111871	0.81950211
5	-0.46536979	0.59152805
6	-0.75541503	0.41360056
7	-0.95911662	0.18380470
8	-1.28155157	-0.09413741
9	-1.58705583	-0.53040186
10	-1.89318453	-1.00999017

```
Faces.dp <- apply(Faces.zROC[-1, ], 1, diff) # similar d' estimates
```

	1	2	3	4	5	6	7	8
	1.0056105	1.0731829	1.0682566	1.0406208	1.0568978	1.1690156	1.1429213	1.1874142
	9	10						

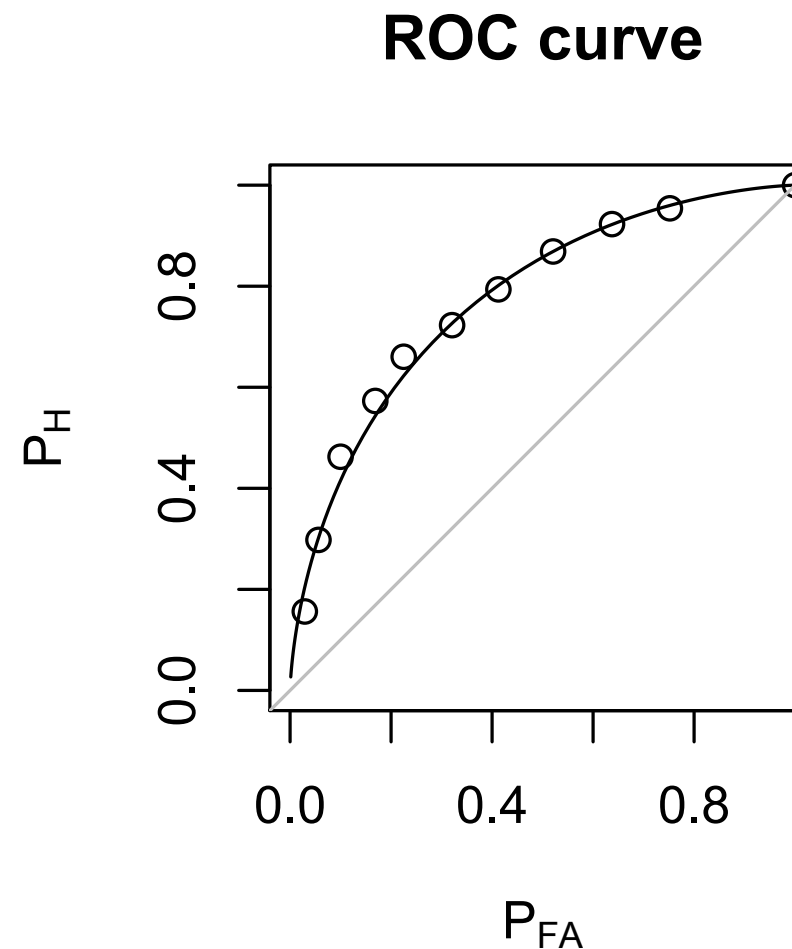
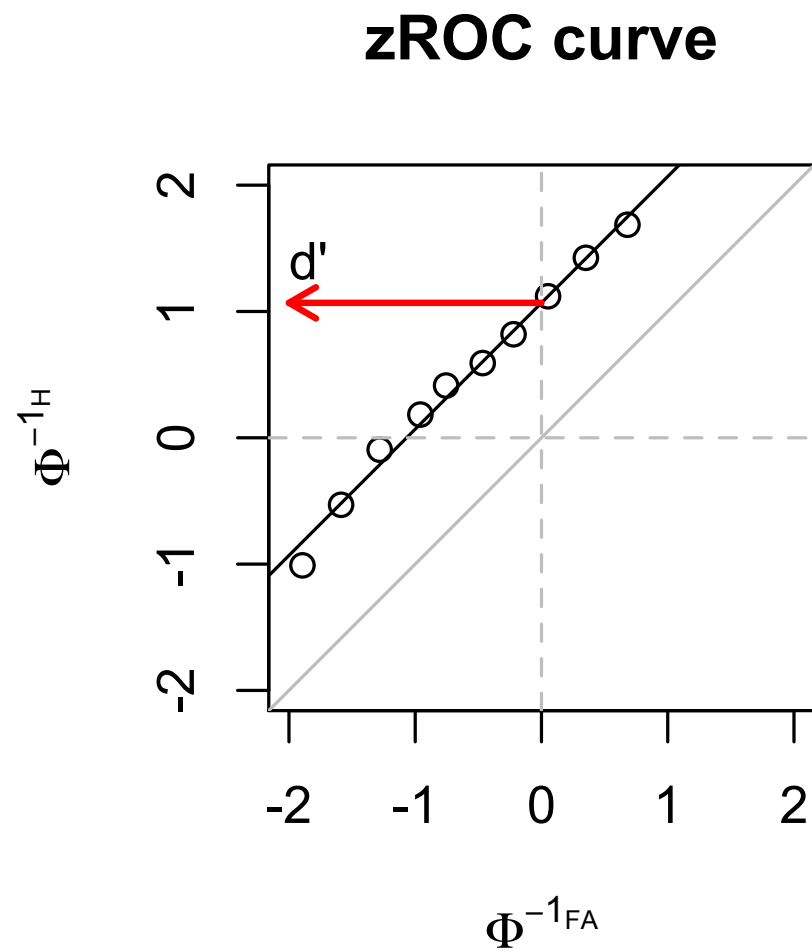
```
1.0566540 0.8831944 # similar d' estimates for each criterion
```

```
c(dp = mean(Faces.dp), SDdp = sd(Faces.dp))
```

dp	SDdp
1.06837681	0.08765449

# Analysis of rating scales: ROC curve

1. Plot zROC curve (H and FA rates transformed by  $\Phi^{-1}$ ), leaving out infinite values
2. Fit a line of unit slope (y-intercept is  $d'$  estimate)
3. Back-transform line to ROC plot using  $\Phi$ .



Only valid when unit slope is acceptable hypothesis. Cannot use least squares fit when slope not equal to 1 because error on both axes.

# Analysis of rating scales: Cumulative Link Models

$$g(\mathbf{E}[Pr(Y < \theta_k \mid x_k)]) = \theta_k - X\beta$$

such that

$$\theta_1 < \theta_2 < \dots < \theta_p$$

$g$  is link function, e.g.,  $\Phi$  for probit case.

Expected probability that rating is less than a criterion value given an internal response,  $x_k$ , is related to a linear predictor.

Multiple ordered intercepts,  $\theta_k$ ,  
and a linear predictor,  $X\beta$

These models can be fit by ordinal regression methods  
(*polr* function in **MASS** package and *clm* function in **ordinal** package in R)

# Analysis of rating scales: Cumulative Link Models

```
library(ordinal)
m1 <- clm(ordered(SimRating) ~ sibs, data = Faces, link = "probit")
summary(m1)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
sibs1	1.07366	0.07023	15.29	<2e-16 ***	# d' estimate similar to previous value

Threshold coefficients:

	Estimate	Std. Error	z value	
0 1	-0.66655	0.05791	-11.510	# criterion estimates
1 2	-0.35394	0.05463	-6.479	
2 3	-0.05270	0.05341	-0.987	
3 4	0.23123	0.05353	4.320	
4 5	0.46783	0.05440	8.600	
5 6	0.69942	0.05602	12.484	
6 7	0.91733	0.05813	15.782	
7 8	1.20936	0.06165	19.616	
8 9	1.60917	0.06771	23.767	
9 10	2.04990	0.07669	26.728	

# Analysis of rating scales: Cumulative Link Models

$$g(E[Pr(Y < \theta_k)]) = \frac{\theta_k - X\beta}{\sigma_i}$$

```
m2 <- clm(ordered(SimRating) ~ sibs, scale = ~sibs, # scale dependent - UVG SDT
          data = Faces, link = "probit")
summary(m2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
sibs1	1.07395	0.07692	13.96	<2e-16 ***

log-scale coefficients:

	Estimate	Std. Error	z value	Pr(> z )
sibs1	0.0005636	0.0602593	0.009	0.993

# log( $\sigma$ ) est. close to 1, eq. var.

```
anova(m1, m2)
```

Likelihood ratio tests of cumulative link models:

	formula:	scale:	link:	threshold:
m1	ordered(SimRating) ~ sibs	~1	probit	flexible
m2	ordered(SimRating) ~ sibs	~sibs	probit	flexible

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
m1	11	4349.7	-2163.8			
m2	12	4351.7	-2163.8	1e-04	1	0.9925 # Eq. Var. Model not rejected!



# Mixed-effects models

Up to here, we have been modeling data from individual observers or simply aggregating all observer data together. But, how do we treat observer differences in SDT models. One approach is to fit each observer individually, extract the parameters of interest (e.g.,  $d'$ ) and fit a linear mixed-effects model to the parameters. This 2-step procedure is referred to as the *Parameters as Outcome Model* (PAOM) and in some hard to fit models is the only tractable method.

But, it is also possible to extend the mixed-effects model to the GLM framework to obtain Generalized Linear Mixed-effects Models (GLMM) that permit estimation of additional sources of variance (e.g., due to observers, stimuli, etc.)

Some references:

DeCarlo, L. T. (2010). On the statistical and theoretical basis of signal detection theory and extensions: Unequal variance, random coefficient, and mixture models. *Journal of Mathematical Psychology*, **54**, 304-313.

Moscatelli, A, Mezzetti, M, Lacquaniti, F (**2012**). Modeling psychophysical data at the population-level: the generalized linear mixed model. *J Vis*, **12**, 11.

Rouder J. N., Lu J., Sun D., Speckman P., Morey R., Naveh-Benjamin M. (2007). Signal detection models with random participant and item effects. *Psychometrika* , **72**, 621–642.

Knoblauch, K., Maloney L.T., *Modeling Psychophysical Data in R*. Springer, New York (2012), Chapter 9

# Linear Mixed-effects Models: lightning review

$$E(Y) = \mathbf{X}\beta + \mathbf{Z}b + \epsilon$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad b \sim \mathcal{N}(0, \Sigma^2)$$

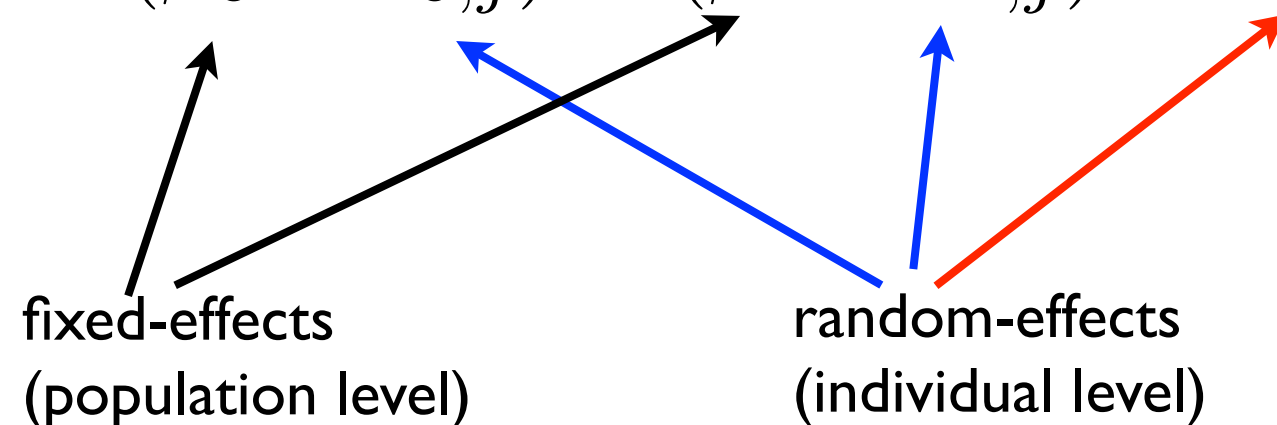
$\mathbf{Z}$  model matrix for random effects

$\Sigma^2$  variance-covariance matrix for vector  $b$

We estimate variances, elements of var-cov matrix, not elements of  $b$

For example,

$$y_{ij} = (\beta_0 + b_{0,j}) + (\beta_1 + b_{1,j})x + \epsilon_{i,j}$$



# Generalized Linear Mixed-effects Models (GLMM)

Bernoulli/Binomial case:

$$g(\mathbf{E}[Pr(Y = 1)] \mid b_i) = \mathbf{X}\beta + \mathbf{Z}b$$

$$b \sim \mathcal{N}(0, \Sigma^2)$$

$g$  is the link function (as before)

Note i) that the response is conditional on the random effects,  
ii) that the variance is known for a binomial model so  $b$   
comprises all of the random effects

# Generalized Linear Mixed-effects Models (GLMM)

```
load("Faces2.Rdata")
str(Faces2)
'data.frame': 960 obs. of 4 variables:
 $ Resp : Factor w/ 2 levels "0","1": 2 2 2 1 2 2 2 2 2 2 ...
 $ Stim : Factor w/ 2 levels "A","P": 2 2 2 2 2 2 2 2 2 2 ...
 $ Obs  : Factor w/ 32 levels "S1","S2","S3",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Image: Factor w/ 30 levels "Im1","Im2","Im3",...: 1 2 3 4 5 6 7 8 9 10 ...
```

From same study as the Faces data set but task was only to judge whether pair of images were siblings or not (no rating, so just a Yes/No experiment).

32 observers - random effect

30 images - images also random because a sample from the set of all images of siblings

Because all 32 observers saw all 30 images (in random order), the random effects are crossed



Are these siblings or not?

Maloney, L. T., and Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision*, 6(10):4, 1047–1056.

# Generalized Linear Mixed-effects Models (GLMM)

Suppose that we ignore random effects of Obs and Image:

```
load("Faces2.Rdata")  
str(Faces2)
```

```
glm(Resp ~ Stim, binomial(probit), Faces2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.35735	0.05855	-6.103	1.04e-09 ***	
StimP	0.99941	0.08506	11.749	< 2e-16 ***	# d' about 1 as reported

# Generalized Linear Mixed-effects Models (GLMM)

$$g(\mathbb{E}[y_{ijk} \mid b_{0j}, b_{0k}]) = (\beta_0 + b_{0j} + b_{0k}) + \beta_1 x_{ijk}$$

Random effects of:                      Obs                      Image   on criterion,  $\beta_0$

```
gm1 <- glmer(Resp ~ Stim + (1 | Obs) + (1 | Image), Faces2, binomial(probit))
summary(gm1)
```

...

Random effects:

Groups	Name	Variance	Std.Dev.	
Obs	(Intercept)	0.01303	0.1142	# $b_{0j}$
Image	(Intercept)	0.55453	0.7447	# $b_{0k}$

Number of obs: 960, groups: Obs, 32; Image, 30

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.4777	0.2051	-2.330	0.0198	*
StimP	1.2573	0.2888	4.354	1.34e-05	*** # higher d' than w/ glm

...

Correlation of Fixed Effects:

	(Intr)	
StimP	-0.703	# high correlation of fixed effects # Can fix by centering explanatory variables

# Generalized Linear Mixed-effects Models (GLMM)

centering explanatory variables and correlation among fixed effects

```
Faces2$LStim <- with(Faces2, ifelse(Stim == "P", 0.5, -0.5))
gm1c <- glmer(Resp ~ LStim + (1 | Obs) + (1 | Image), Faces2, binomial(probit))
summary(gm1c)
```

...

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

Obs	(Intercept)	0.01303	0.1142	# Random effects unaffected
-----	-------------	---------	--------	-----------------------------

Image	(Intercept)	0.55452	0.7447
-------	-------------	---------	--------

Number of obs: 960, groups: Obs, 32; Image, 30

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.1509	0.1458	1.035	0.301	# $\beta_0$ shifted
LStim	1.2573	0.2888	4.354	1.34e-05 ***	# $\beta_1$ unaffected

...

Correlation of Fixed Effects:

(Intr)

LStim	0.001	# correlation reduced
-------	-------	-----------------------

# Generalized Linear Mixed-effects Models (GLMM)

Other random effects fitted and tested - nested sequences of models

$$g(E[y_{ijk} | b_{0j}, b_{0k}]) = (\beta_0 + b_{0j} + b_{0k}) + (\beta_1 + b_{1j} + b_{1k})x_{ijk}$$

```
gm2a <- glmer(Resp ~ LStim + (LStim + 0 | Obs) +          # rx sensitivity x Obs
  (1 | Obs) + (1 | Image), Faces2, binomial(probit))
gm2b <- glmer(Resp ~ LStim + (1 | Obs) + (1 | Image) +
  (LStim + 0 | Image), Faces2, binomial(probit))      # rx sensitivity x Image
gm3 <- glmer(Resp ~ LStim + (LStim + 0 | Obs) + (LStim + 0 | Image) +
  (1 | Obs) + (1 | Image), Faces2, binomial(probit)) # All rx's combined
```

```
anova(gm1c, gm2a, gm3) #LR test of one sequence
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
gm1c	4	1015.9	1035.3	-503.93	1007.9				
gm2a	5	1015.5	1039.8	-502.73	1005.5	2.4174		1	0.1200
gm3	6	1017.5	1046.7	-502.73	1005.5	0.0000		1	0.9988

```
anova(gm1c, gm2b, gm3) #LR test of other sequence
```

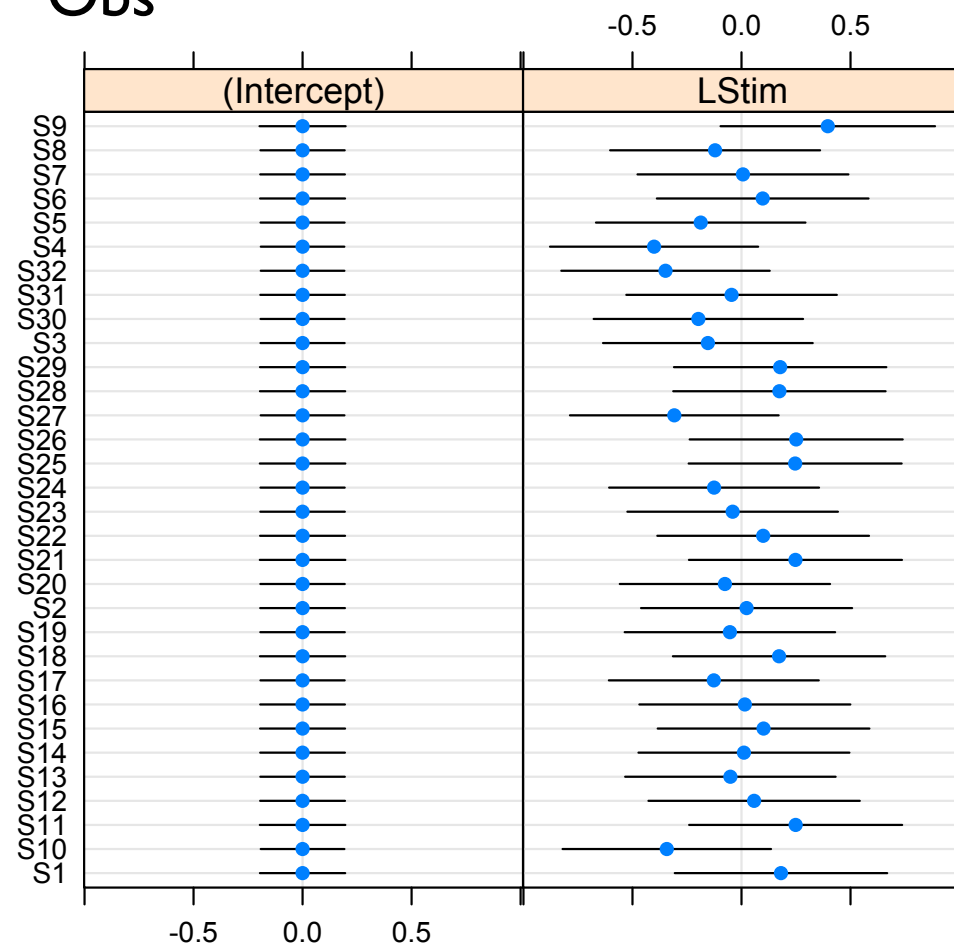
	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
gm1c	4	1015.9	1035.3	-503.93	1007.9				
gm2b	5	1017.9	1042.2	-503.93	1007.9	0.0000		1	1.00
gm3	6	1017.5	1046.7	-502.73	1005.5	2.4174		1	0.12

```
# Test is conservative (forcing variance to border of parameter space)
# Could reduce p-value by 1/2 or do a bootstrap test (more accurate).
# Suggests that p ~ 0.1, close to significance
```

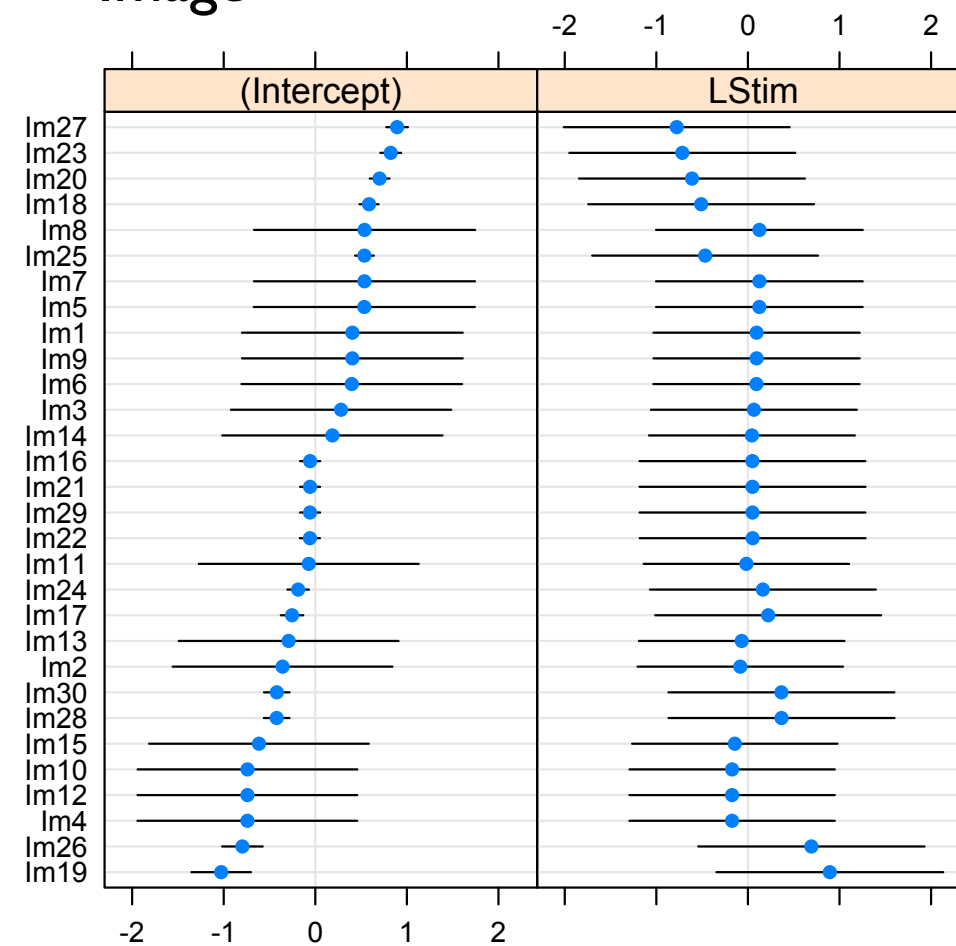


# Caterpillar plots of random effects

Obs



Image



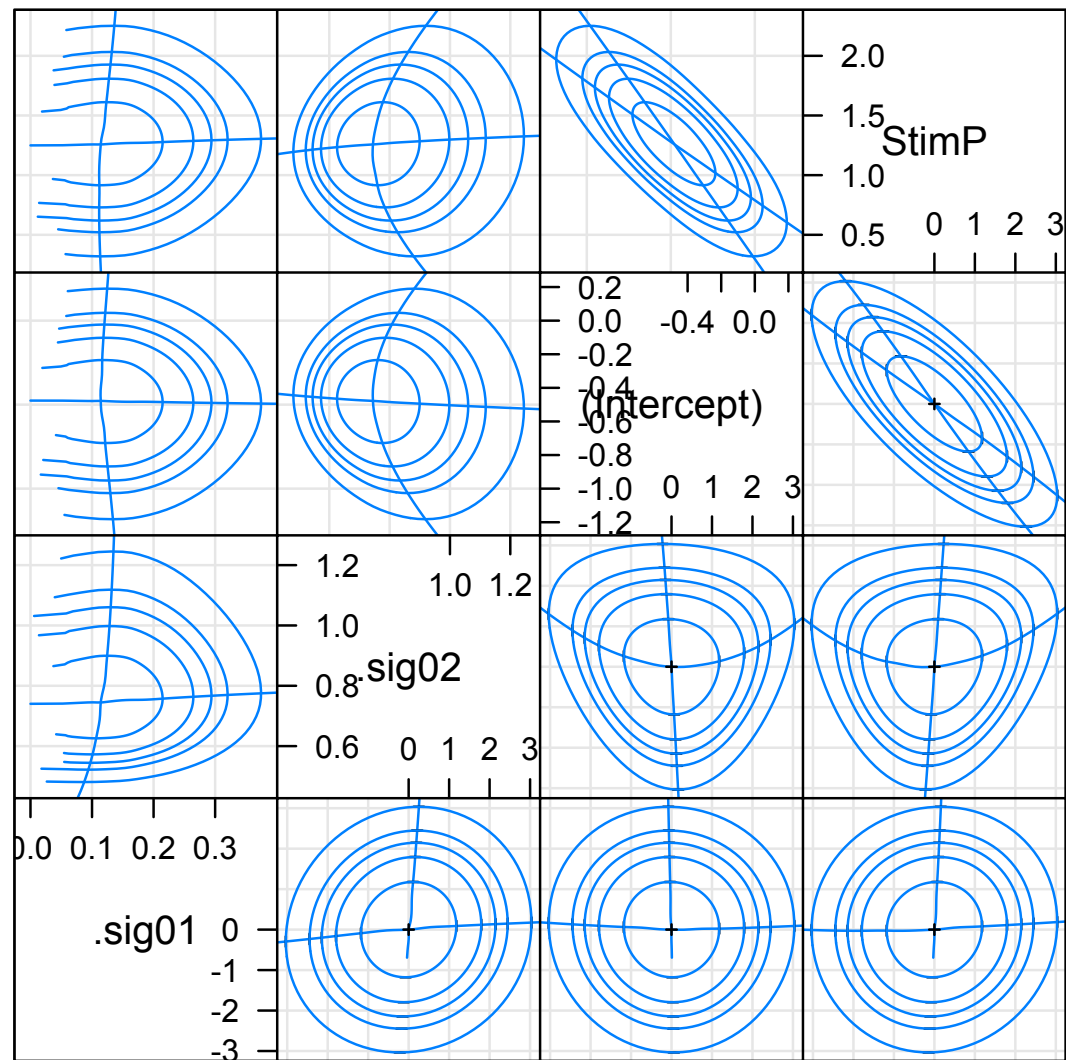
Random effects:

Groups	Name	Variance	Std.Dev.
Obs	(Intercept)	0.01401	0.1184
Obs	LStim	0.12975	0.3602
Image	(Intercept)	0.05438	0.2332
Image	LStim	2.08581	1.4442

Not quite working correctly for this version of software, but can be informative about sources of variance. Some images generate almost perfect categorization, others more difficult. Some observers almost perfect responses. Model would probably fit better with more responses per observer.

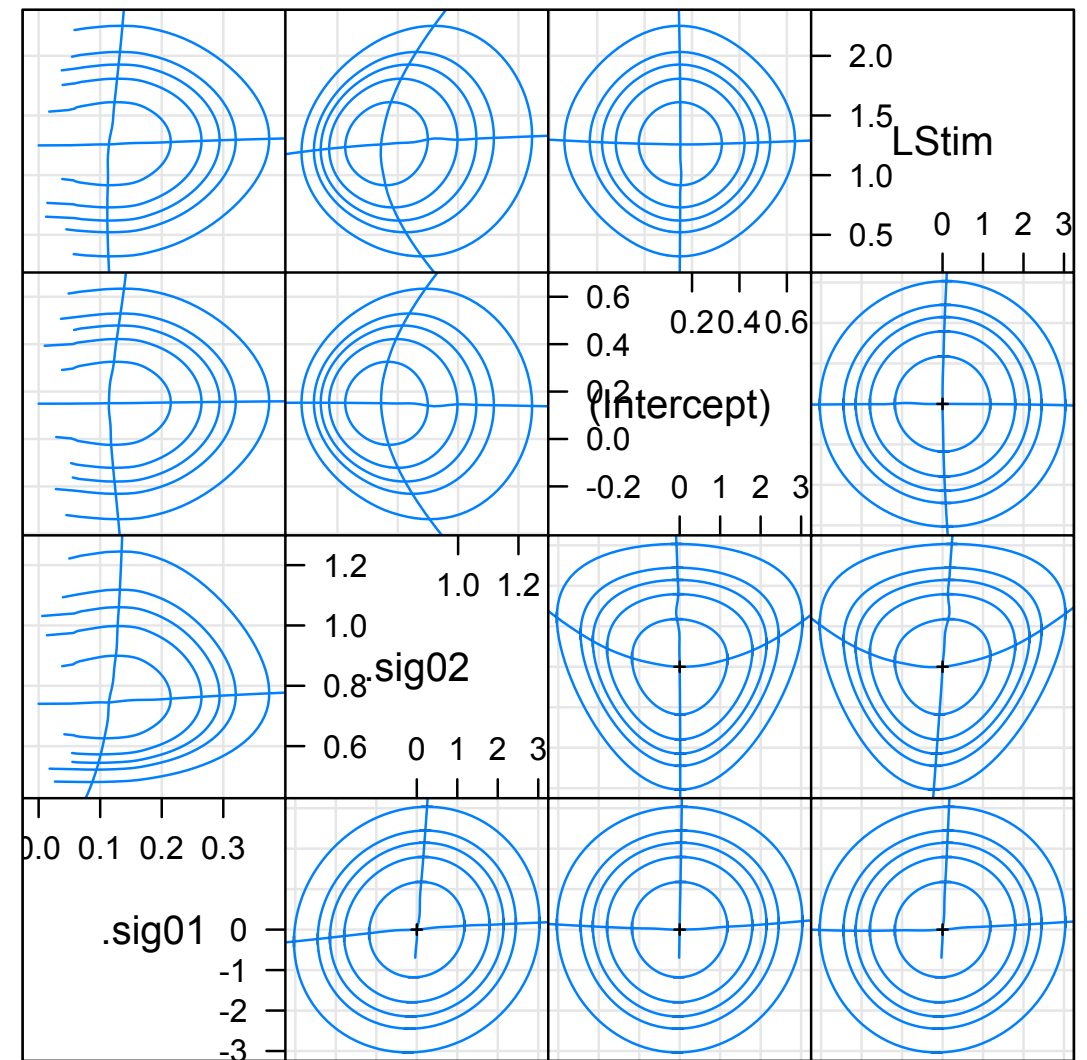
# Profile plots of likelihood surface

correlated fixed effects



Scatter Plot Matrix

uncorrelated fixed effects



Scatter Plot Matrix

Plots above diagonal for parameters; plots below for log(parameters).  
Note how contours for variances do not close at boundary above diagonal  
and are asymmetric,  
but below log(variance) profiles are more well-behaved

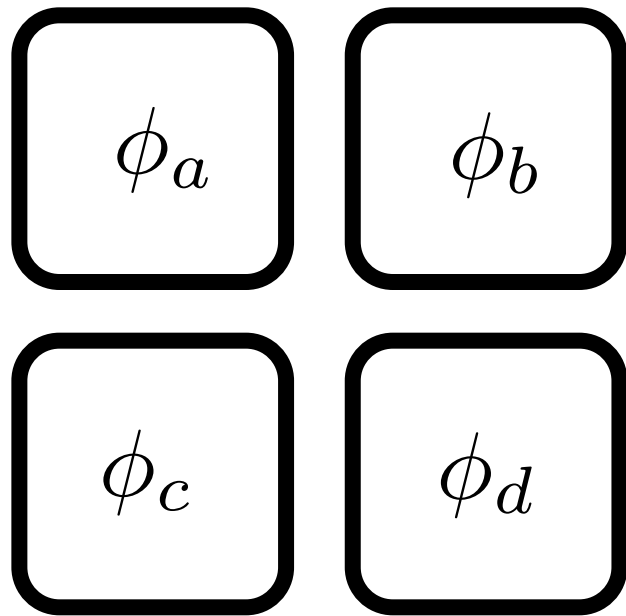
# Extension of SDT to scaling: MLDS

MLDS: Maximum Likelihood Difference Scaling

stimulus dimension  $\phi_1$   $\phi_2$   $\dots$   $\phi_p$



Quadruples



Signal Detection Model

Decision rule:

$$\Delta_i = (\psi_b - \psi_a) - (\psi_d - \psi_c) + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

Choose upper pair if  $\Delta_i > 0$ ,  
otherwise choose lower pair

Task: Between which pair (a, b) or (c, d) is the difference greatest?

Estimate scale values,  $\psi_1, \psi_2, \dots, \psi_p$ , by maximum likelihood that best predict the observer's choices. Estimated scale has interval properties. Equal scale differences correspond to equal perceptual differences.

Maloney & Yang (2003) *J. Vision*, **3**, 573-585

Knoblauch & Maloney (2010) *J. Stat. Software*, **25**, 1-25

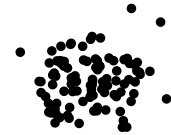
Knoblauch & Maloney (2012) **Modeling Psychophysical Data in R**, Springer

# Example: Correlation in random scatterplots

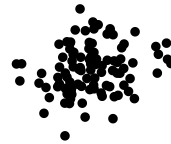
$r = 0$



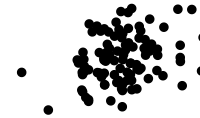
$r = 0.1$



$r = 0.2$



$r = 0.3$



$r = 0.4$



$r = 0.5$



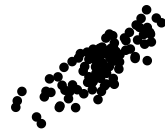
$r = 0.6$



$r = 0.7$



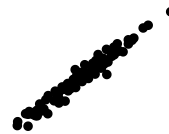
$r = 0.8$



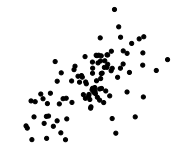
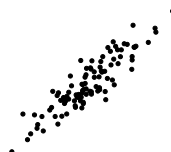
$r = 0.9$



$r = 0.98$



Typical trial



Do you perceive the difference in correlations to be greater between the top or the bottom pair?

# Extension of SDT to scaling: MLDS

Decision rule:

$$\Delta_i = (\psi_b - \psi_a) - (\psi_d - \psi_c) + \epsilon$$
$$\epsilon \sim N(0, \sigma^2)$$

Choose upper pair if  $\Delta_i > 0$ ,  
otherwise choose lower pair

Likelihood:

$$\mathcal{L}(\Psi, \sigma; R_i) = \prod_{i=1}^n \Phi\left(\frac{\Delta_i}{2\sigma}\right)^{R_i} \left(1 - \Phi\left(\frac{\Delta_i}{2\sigma}\right)\right)^{1-R_i},$$

an *equal-variance Gaussian model*, scaled so that  $\psi_i$  are in units of ***d'*** (in theory!).

# Extension of SDT to scaling: MLDS

Decision rule is linear combination of internal responses to stimuli

$$\Delta_i = (\psi_b - \psi_a) - (\psi_d - \psi_c) + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

Choose upper pair if  $\Delta_i > 0$ ,  
otherwise choose lower pair

So we can reformulate the model as a binomial GLM:

$$g(\mathbf{E}[Pr(R = 1)]) = \psi_2 x_2 + \cdots + \psi_p x_p$$

$$\Phi^{-1}(Pr(Y = 1)) = \mathbf{X} \beta \quad \beta = (\psi_2, \dots, \psi_p)$$

Example, first six lines of  $X$ , 1 column for each stimulus, 1 row for each trial

	stim.2	stim.3	stim.4	stim.5	stim.6	stim.7	stim.8	stim.9	stim.10	stim.11
1	0	-1	0	-1	0	1	0	0	0	
0	0	1	0	1	0	0	-1	0	0	
-1	1	0	0	0	0	1	0	-1	0	0
-1	0	0	1	1	-1	0	0	0	0	0
0	1	0	0	0	1	-1	0	0	0	0
0	0	0	1	1	0	0	-1	0	0	0
					...					

# Extension of SDT to scaling: MLDS

Can fit model matrix to responses (I/O, i.e., upper/lower pair choices) by either direct ML estimation or via GLM. **MLDS** package implements both methods.

```
library(MLDS)
```

```
kk.mlds <- mlds(kk1) # data from estimation of correlation in scatterplots
```

Perceptual Scale:

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98
0.000	-0.217	0.173	-0.202	0.421	1.374	1.889	2.453	3.040	3.982	5.440

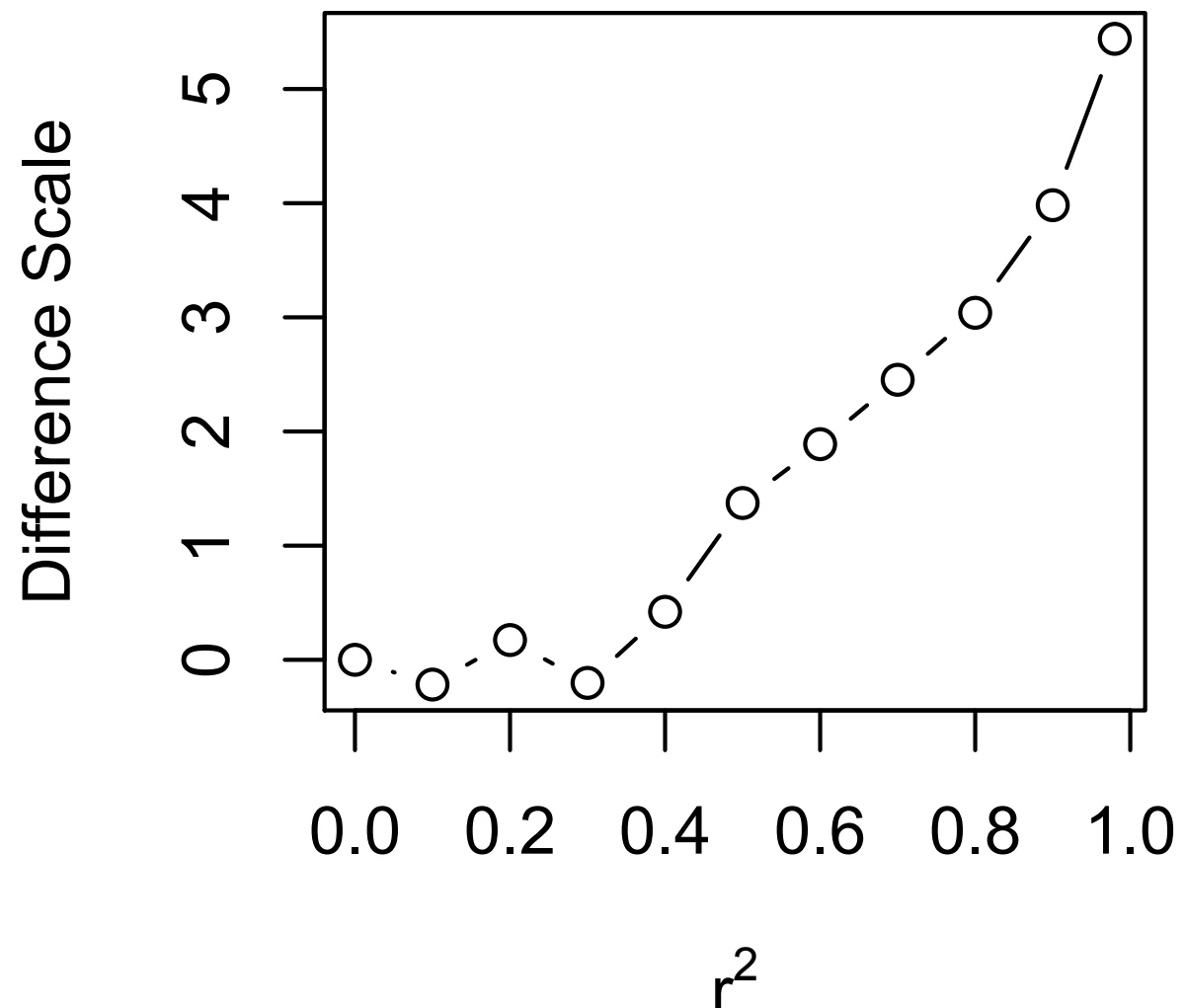
sigma:

```
[1] 1
```

```
plot(kk.mlds, type = "b")
```

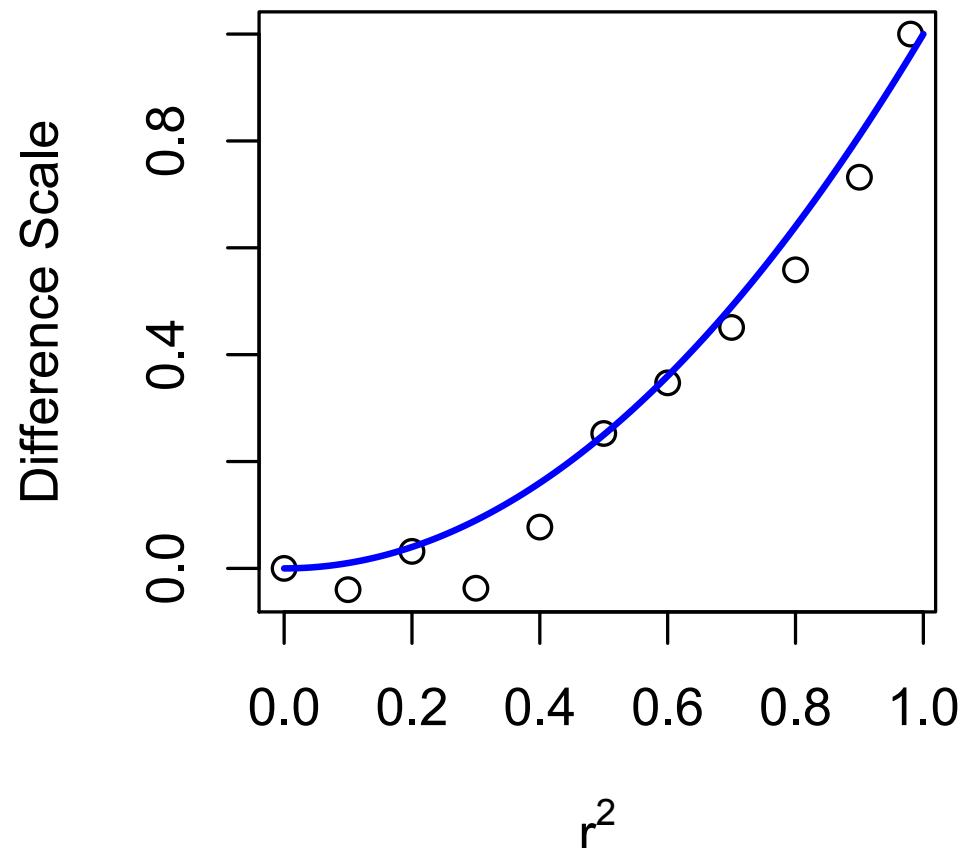
Looks a bit like perceived differences in correlation follow  $r^2$ .

As scaled, ordinate values correspond to  $d'/2$ , so need to multiply by 2 to put in units of  $d'$ .



# Extension of SDT to scaling: MLDS

```
plot(kk.mlds, standard.scale = TRUE, # normalized to (0, 1) as in Maloney & Yang (2003)
     xlab = expression(r^2), ylab = "Difference Scale")
xx <- seq(0, 1, len = 100)
lines(xx, xx^2, lwd = 2, col = "blue")
```



```
# fitting with the formula method
# initial estimate of p and sigma
mls(~ x^p, p = c(2, 0.1), data = kk1)$par

[1] 2.161583 # pretty close to 2
```

It is possible to apply mixed-effects models to MLDS.

See Knoblauch & Maloney (2012) *Modeling Psychophysical Data in R*, Springer for further details.



## Extensions of SDT: perspectives

Any time that you can specify the decision rule for a choice experiment, you ought to be able to estimate the parameters by maximum likelihood.

If you can express the decision rule as a linear predictor, then you ought to be able to express the model as a GLM and fit it with off-the-shelf GLM methods, like `glm` in R.

See Knoblauch & Maloney (2012) *Modeling Psychophysical Data in R*, Springer for further details and applications of `glm` to fitting psychophysical data SDT framework.

Many of these methods easily extended to Bayesian framework, but that would be the subject of a separate course.

