More Latent Models

SEM as one of a family of latent models

I. Data = Model + Error

II. sems are models of covariance structures

   A. typical sems model covariaces

   B. change models are models of moments and include means

III. Several alternative latent models

Alternative latent models

I. Item Response Theory (IRT)

   A. 1, 2 and 3 parameter models

II. Latent Class Analysis (LCA)

CTT vs. IRT

I. Classical Test Theory as a covariance-structure model

II. IRT as a data model

Classical Test Theory

I. Classical test theory is a model of covariances

   A. items are sampled from larger domain

   B. items are random replicates of each other

   C. difficulty of the item is not included in the model

Classical Test Theory

I. X as data matrix with elements x_{ij}

   A. X is N subjects by k items $_{N}X_{k}$

II. Covariance of items is X'X/N = C_{xx}

III. Item covariances reflect domain or true score

IV. Item variances reflect domain + specific + error scores

V. Reliability of a test X is the amount of true score in the test

Classical Reliabilty

I. $r_{xx} = 1 - \sigma^2_e / \sigma^2_x$ or $(\sigma^2_x - \sigma^2_e)/\sigma^2_x$

A. problem is how to estimate σ^2_e

B. σ^2_e is the sum of error variances for all items

C. Multiple estimates of σ^2_e

1. 1 - average correlation (used for alpha)
2. 1 - squared multiple correlation (Guttman L_6)
3. 1 - communality (McDonald $omega_{total}$)

CTT and reliability

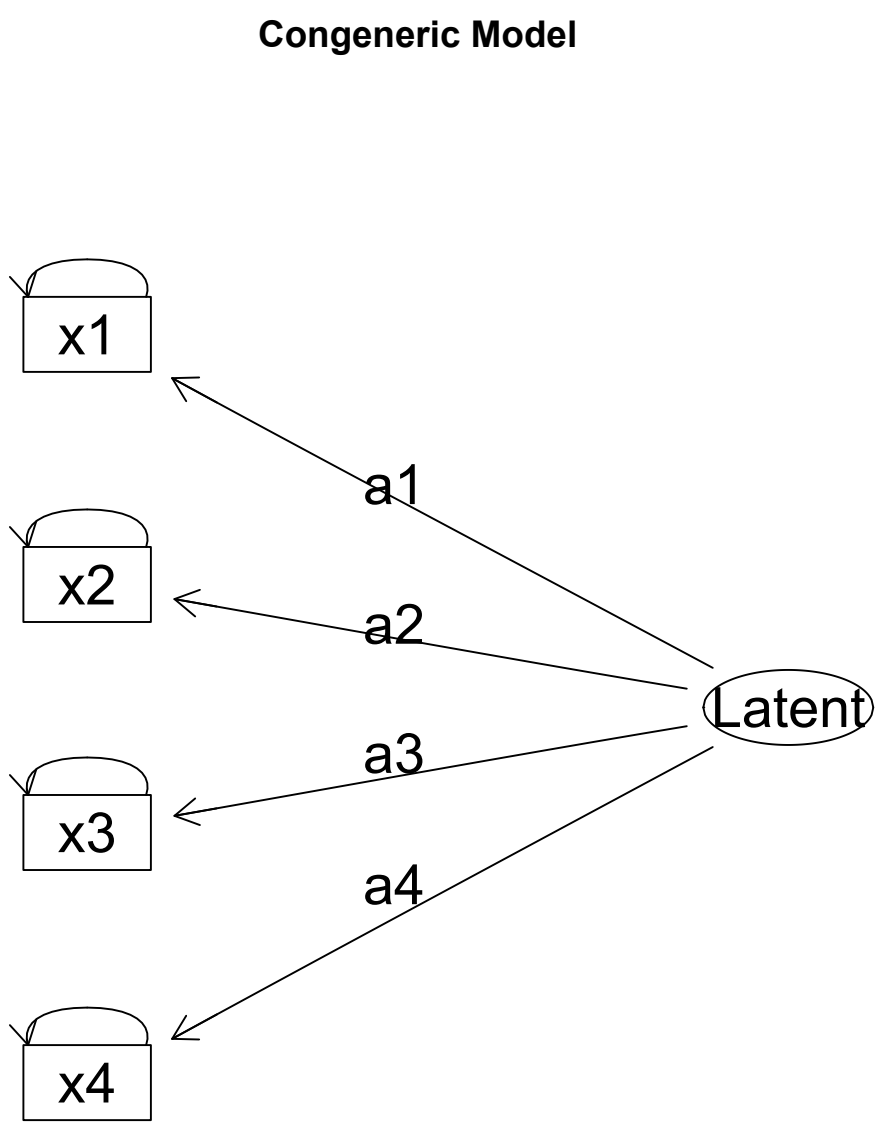
I. Congeneric reliability -- factor model

II. Reliability as the squared correlation with a latent factor

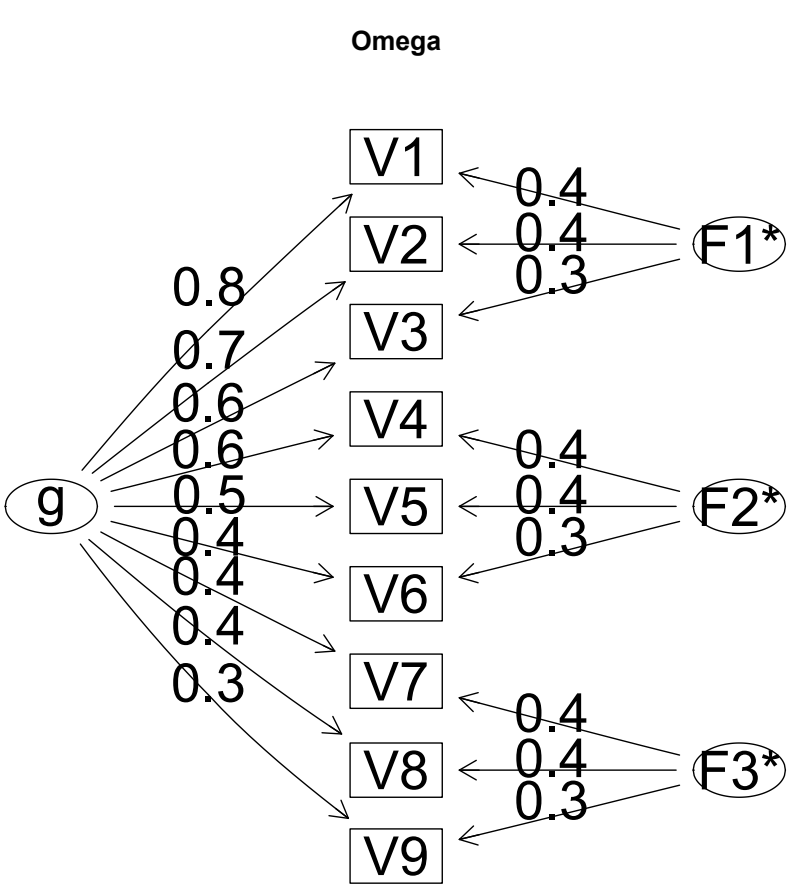
III. Can be estimated if we have at least 3 tests

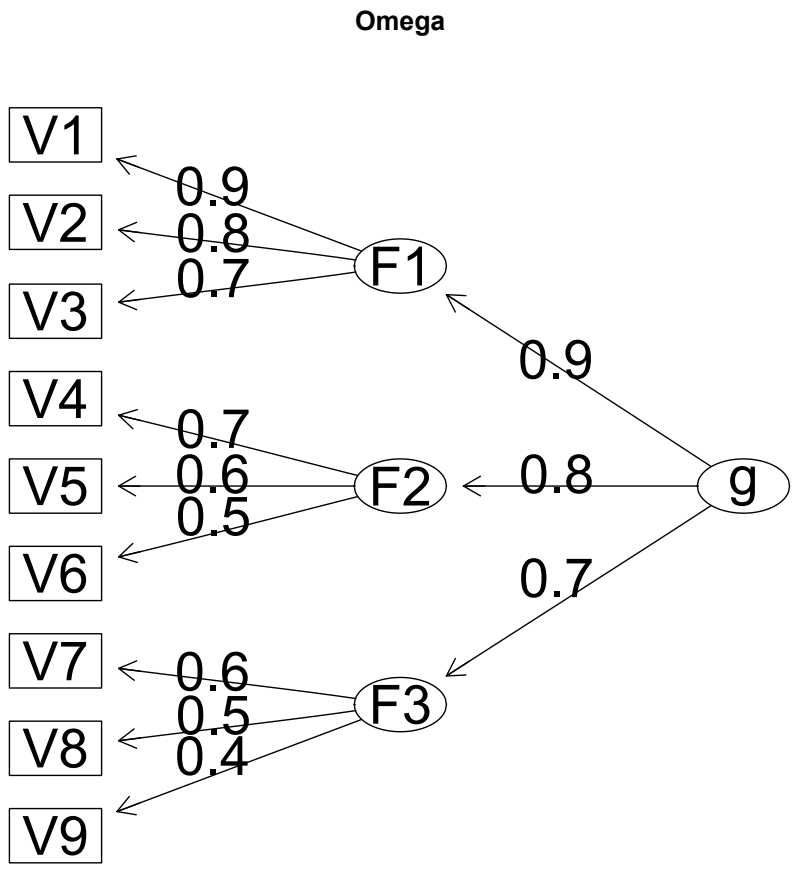
IV. Can be tested if we have four tests

Congeneric test model



Bifactor and hierarchical models





Reliability of hierarchical models

I. $Omega_{hierarchical}$ is σ^2_g / σ^2_x or the amount of general factor saturation

II. $Omega_{total}$ is $(\sigma^2_g+\sigma^2_{f1}+\sigma^2_{f2}+\sigma^2_{f3}) / \sigma^2_x$ or the total common variance divided by the total variance

CTT and reliability

I. Reliability is a sample concept

   A. Sample of people

   B. Sample of items

II. As person variance increases, so will reliability

III. As item homogeneity increases, so will reliability

IV. Does not tell reliability for a single person

Item Response Theory

I. model the response to individual items

II. As a function of person parameter and item parameters

   A. Difficulty

   B. Discriminability

   C. Guessing

The New Psychometrics- Item Response Theory

- Classical theory estimates the correlation of item responses (and sums of items responses, i.e., tests) with domains.
- Classical theory treats items as random replicates but ignores the specific difficulty of the item, nor attempts to estimate the probability of endorsing (passing) a particular item

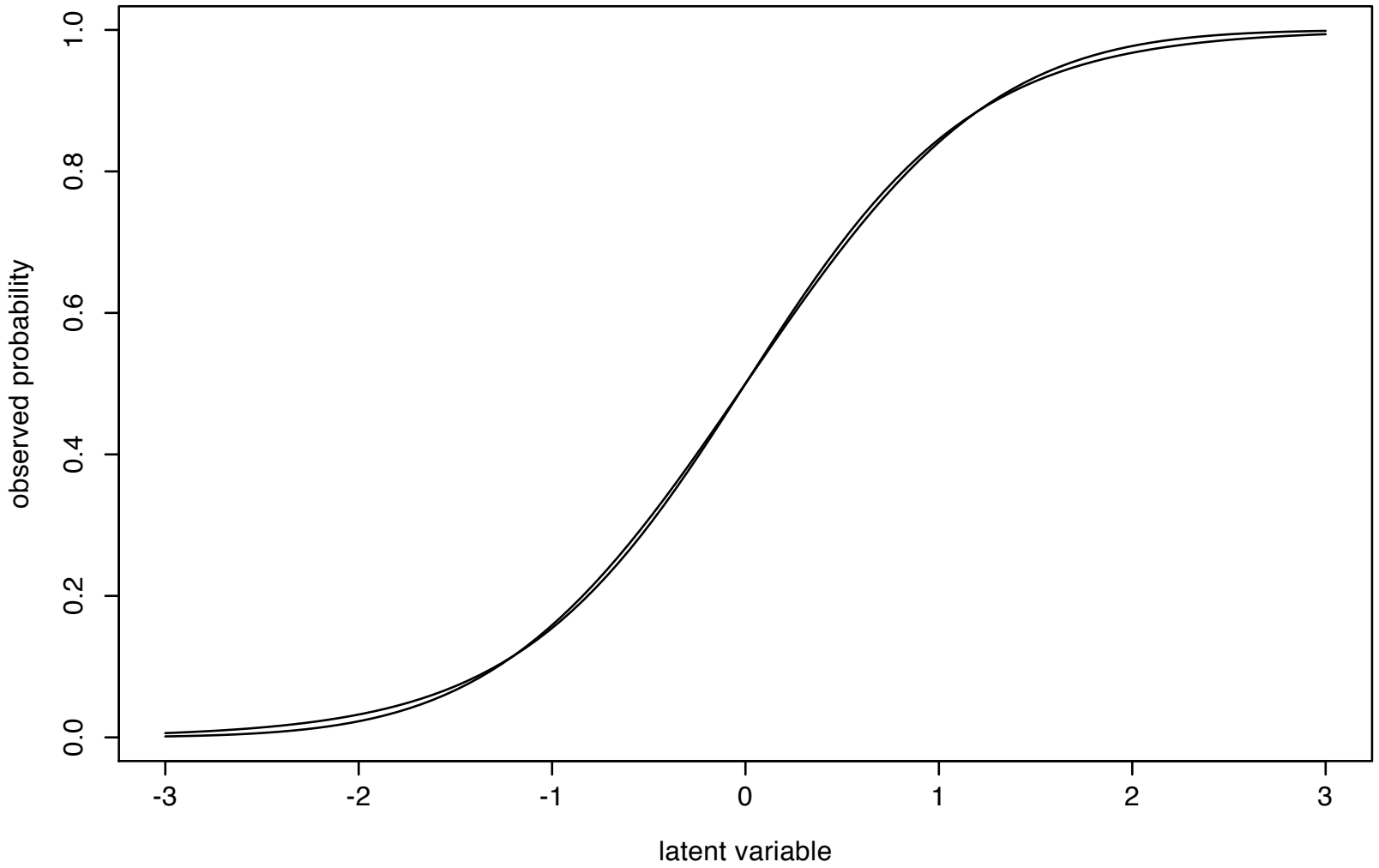
Item Response Theory

- Consider the person's value on an attribute dimension (θ_i).
- Consider an item as having a difficulty δ_j
- Then the probability of endorsing (passing) an item j for person i= $f(\theta_i, \delta_j)$
- $p(\text{correct} \mid \theta_i, \delta_j) = f(\theta_i, \delta_j)$
- What is an appropriate function?
- Should reflect $\delta_j - \theta_i$ and yet be bounded 0,1.

Item Response Theory

- $p(\text{correct} \mid \theta_i, \delta_j) = f(\theta_i, \delta_j) = f(\delta_j - \theta_i)$
- Two logical functions:
  - Cumulative normal (see, e.g., Thurstonian scaling)
  - Logistic $= 1/(1+\exp(\delta_j - \theta_i))$ (the Rasch model)
  - Logistic with weight of 1.7
    - $1/(1+\exp(1.7*(\delta_j - \theta_i)))$ approximates cumulative normal

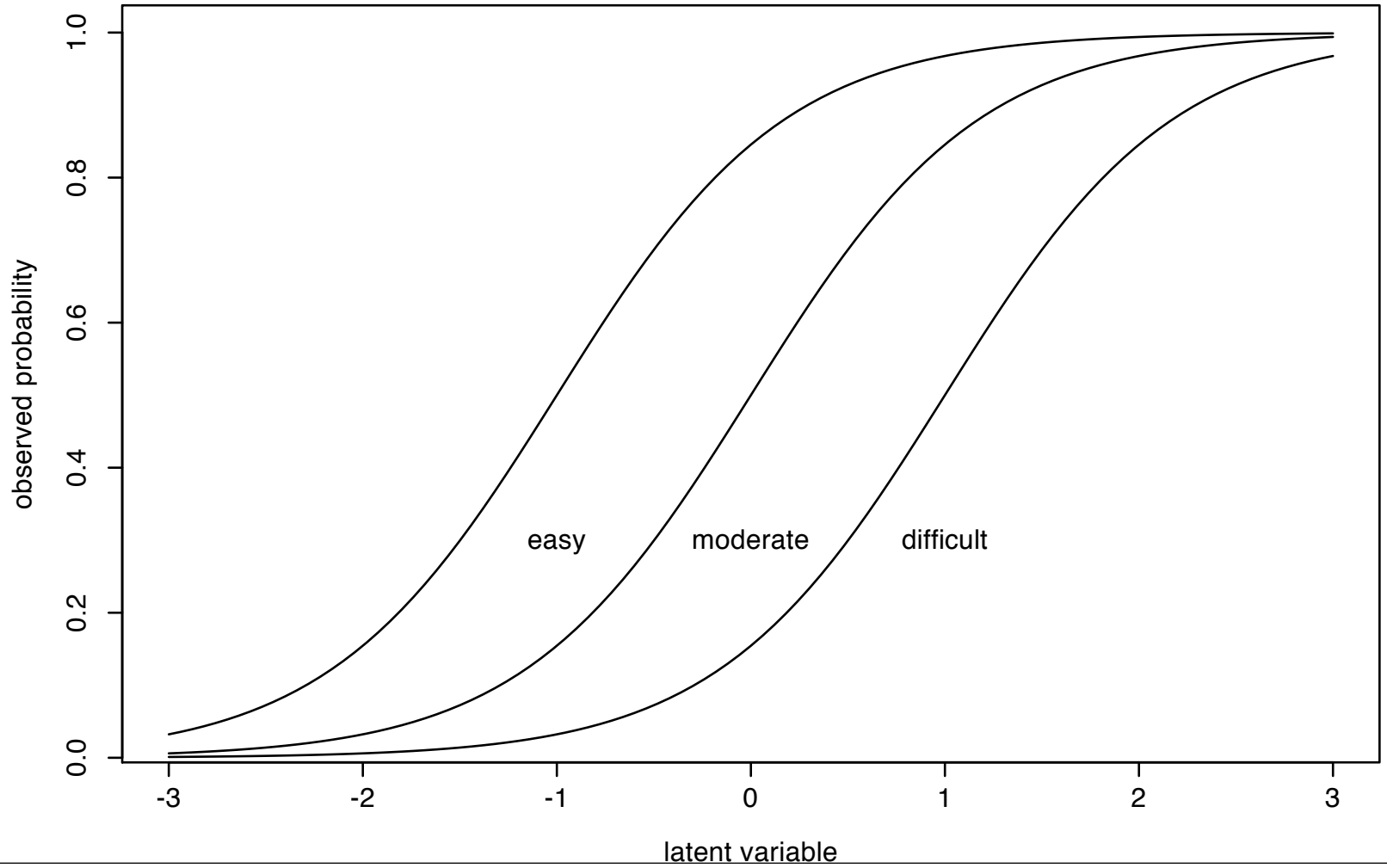
Logistic and cumulative normal



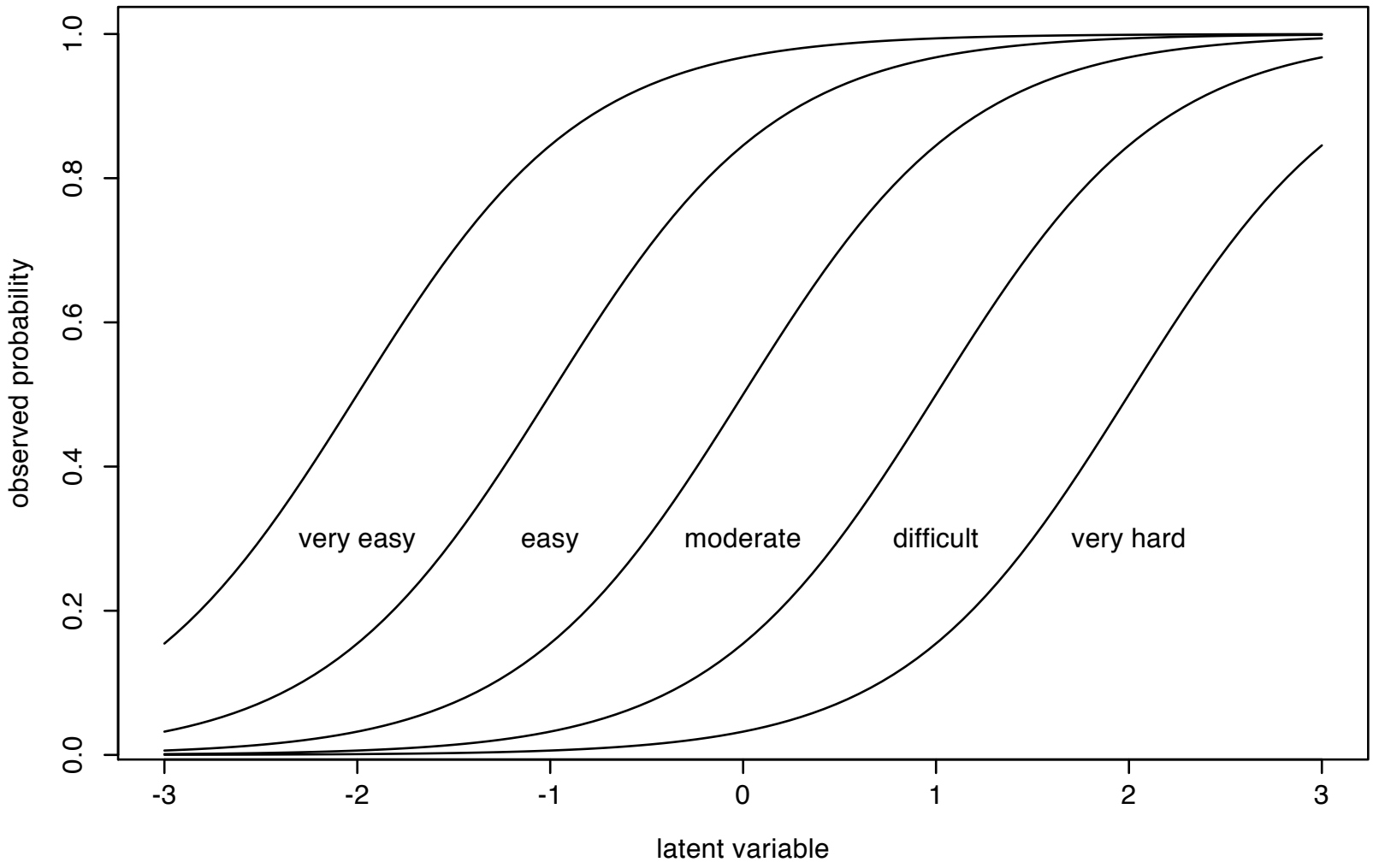
Item difficulty and ability

- Consider the probability of endorsing an item for different levels of ability and for items of different difficulty.
- Easy items ($\delta_j = -1$)
- Moderate items ($\delta_j = 0$)
- Difficulty items ($\delta_j = 1$)

IRT of three item difficulties



item difficulties = -2, -1, 0 , 1, 2



Estimation of ability for a particular person for known item difficulty

- The probability of any pattern of responses (x1, x2, x3, …. Xn) is the product of the probabilities of each response $\prod(p(xi))$.
- Consider the odds ratio of a response
  - $p/(1-p) = 1/(1+\exp(1.7*(\delta_j - \theta_i))) /(1- 1/(1+\exp(1.7*(\delta_j - \theta_i))))$ =
  - $p/(1-p) = \exp(1.7*(\delta_j - \theta_i))$ and therefore:
  - $\text{Ln(odds)} = 1.7* (\theta_i - \delta_j)$ and
  - $\text{Ln (odds of a pattern)} = 1.7\Sigma (\theta_i - \delta_j)$ for known difficulty

Unknown difficulty

- Initial estimate of ability for each subject (based upon total score)
- Initial estimate of difficulty for each item (based upon percent passing)
- Iterative solution to estimate ability and difficulty (with at least one item difficulty fixed.

IRT using R

- Use the ltm package (requires MASS)
- example data sets include LSAT and Abortion attitudes
- Lsat[1:10,] shows some data
- describe(LSAT) (means and sd)
- m1 <- rasch(Lsat)

Consider data from the LSAT

	Item 1	Item 2	Item 3	Item 4	Item 5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	0	0	0	0	1
6	0	0	0	0	1
7	0	0	0	0	1
8	0	0	0	0	1
9	0	0	0	0	1
10	0	0	0	1	0
			...		

Descriptive stats

describe(Lsat)

	n	mean	sd	median	min	max	range	skew	se
Item 1	1000	0.92	0.27	1	0	1	1	-3.20	0.01
Item 2	1000	0.71	0.45	1	0	1	1	-0.92	0.01
Item 3	1000	0.55	0.50	1	0	1	1	-0.21	0.02
Item 4	1000	0.76	0.43	1	0	1	1	-1.24	0.01
Item 5	1000	0.87	0.34	1	0	1	1	-2.20	0.01

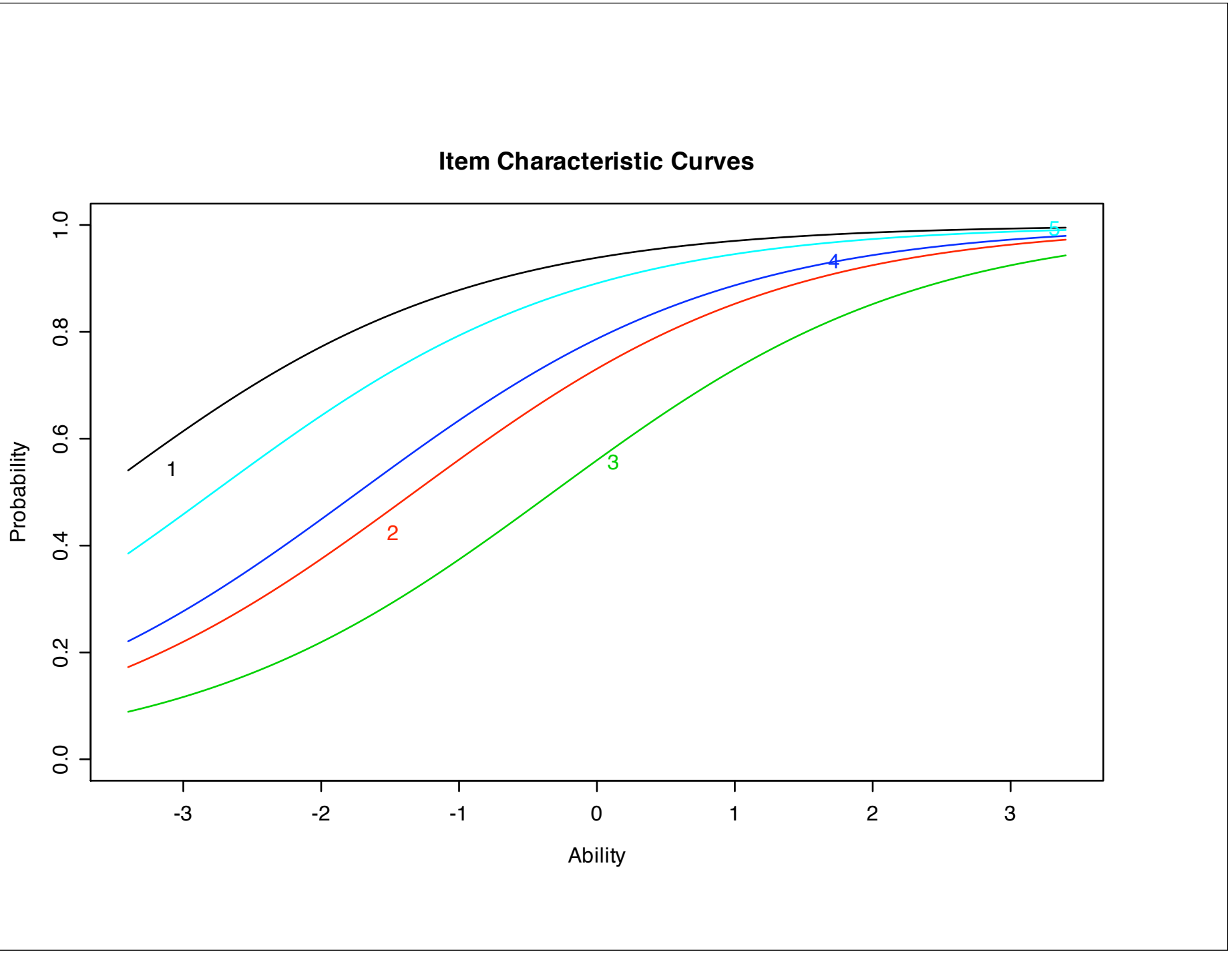
Correlations and alpha

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1.00	0.07	0.10	0.04	0.02
Item 2	0.07	1.00	0.11	0.06	0.09
Item 3	0.10	0.11	1.00	0.11	0.05
Item 4	0.04	0.06	0.11	1.00	0.10
Item 5	0.02	0.09	0.05	0.10	1.00

```
cl <- cor(Lsat)
 Vt <- sum(cl)
iv <- sum(diag(cl))
alpha <- ((Vt-iv)/Vt)*(5/4)
 alpha
[1] 0.29
```

Rasch model

```
m1 <- rasch(Lsat)
 coef(m1,TRUE)
        beta.i  beta P(x=1|z=0)
Item 1   2.730 0.755       0.939
Item 2   0.999 0.755       0.731
Item 3   0.240 0.755       0.560
Item 4   1.306 0.755       0.787
Item 5   2.099 0.755       0.891
```


Item Characteristic Curves
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
Ability
1
2
3
4
5

Classical versus the “new”

- Ability estimates are logistic transform of total score and are thus highly correlated with total scores, so why bother?
- IRT allows for more efficient testing, because items can be tailored to the subject.
- Maximally informative items have p(passing given ability and difficulty) of .5
- With tailored tests, each person can be given items of difficulty appropriate for them.

Computerized adaptive testing

- CAT allows for equal precision at all levels of ability
- CAT/IRT allows for individual confidence intervals for individuals
- Can have more precision at specific cut points (people close to the passing grade for an exam can be measured more precisely than those far (above or below) the passing point.

Psychological (non-psychometric) problems with CAT

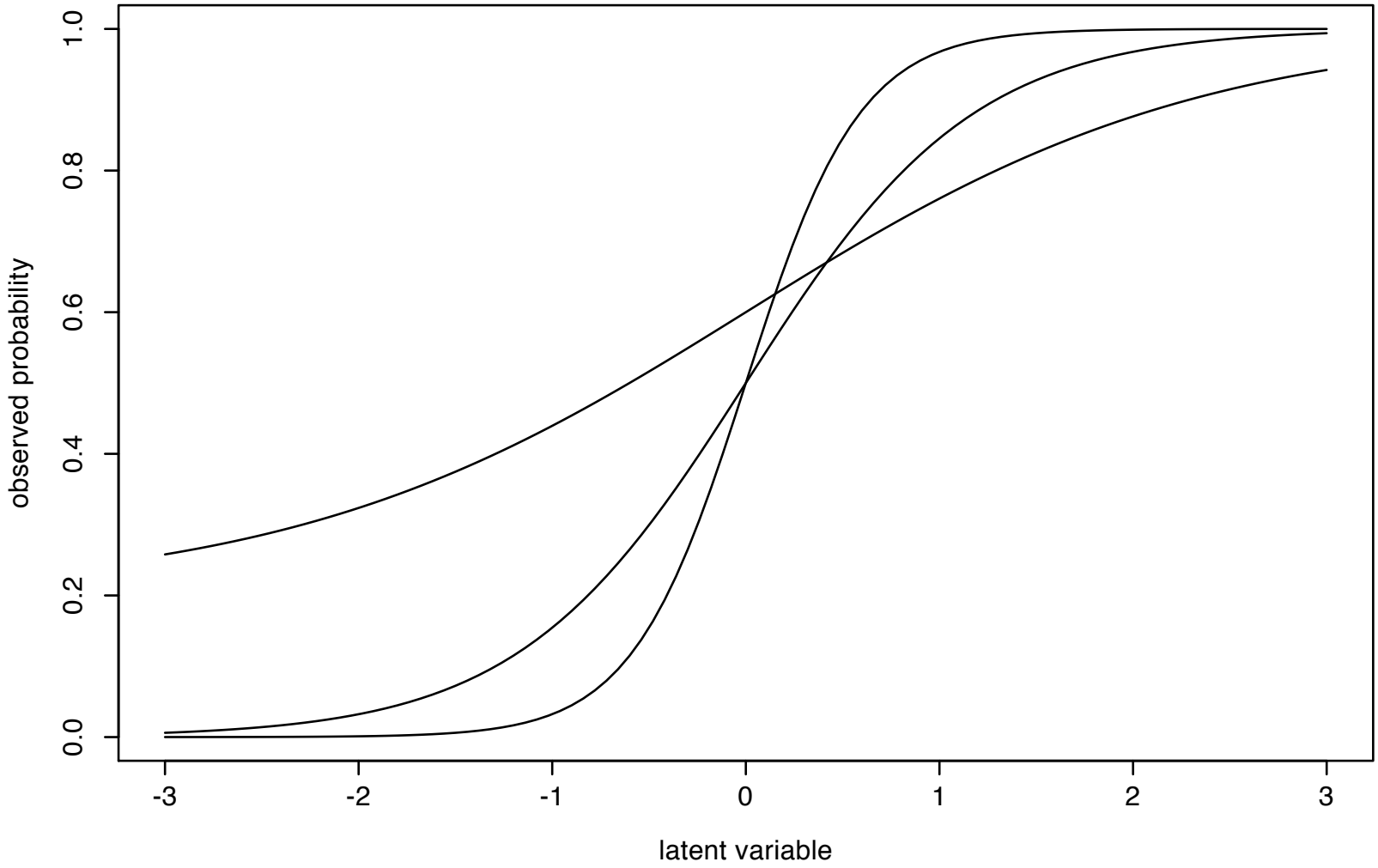
- CAT items have difficulty level tailored to individual so that each person passes about 50% of the items.
- This increases the subjective feeling of failure and interacts with test anxiety
- Anxious people quit after failing and try harder after success -- their pattern on CAT is to do progressively worse as test progresses (Gershon, 199x, in preparation)

Generalizations of IRT to 2 and 3 item parameters

- Item difficulty
- Item discrimination (roughly equivalent to correlation of item with total score)
- Guessing (a problem with multiple choice tests)
- 2 and 3 parameter models are harder to get consistent estimates and results do not necessarily have monotonic relationship with total score

3 parameter IRT

slope, location, guessing



Item Response Theory

- Can be seen as a generalization of classical test theory, for it is possible to estimate the correlations between items given assumptions about the distribution of individuals taking the test
- Allows for expressing scores in terms of probability of passing rather than merely rank orders (or even standard scores). Thus, a 1 sigma difference between groups might be seen as more or less important when we know how this reflects chances of success on an item
- Emphasizes non-linear nature of response scores.

IRT and items

I. Advantage of fitting the raw data rather than the structure.

II. As a sem, it is a latent variable of the moments as well as the means.

III. Nonlinear structure model

Latent Variables in psychopathology

I. Tendency to apply categorical diagnoses

II. But these diagnoses are “comorbid”

III. Can we find a latent model to account for them?

Consider the following matrix of “comorbidity”

	V1	V2	V3	V4
V1	149	48	28	10
V2	48	102	16	6
V3	28	16	61	1
V4	10	6	1	19

Consider the following matrix of "comorbidity"

	V1	V2	V3	V4
V1	149	48	28	10
V2	48	102	16	6
V3	28	16	61	1
V4	10	6	1	19

Diagonal reflect diagnoses, off diagonal, comorbidities
but what are the marginals?

Need to know the marginals!

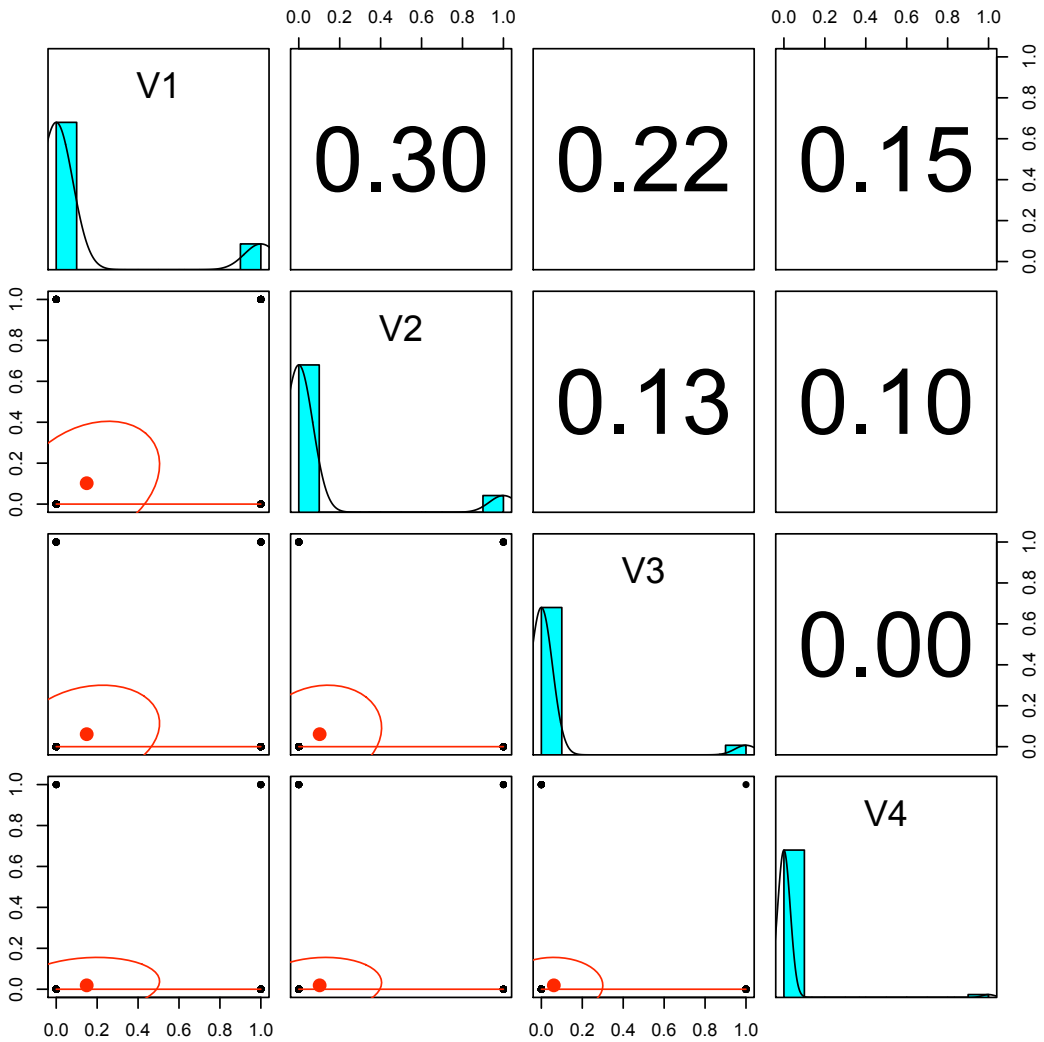
I. Just reporting co-occurrences of two categories is not enough

II. Need to know frequencies of diagnosis and non diagnosis

Data generating "comorbidities"

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
V1	1	1000	0.15	0.36	0	0.06	0	0	1	1	1.97	1.88	0.01
V2	2	1000	0.10	0.30	0	0.00	0	0	1	1	2.63	4.90	0.01
V3	3	1000	0.06	0.24	0	0.00	0	0	1	1	3.66	11.43	0.01
V4	4	1000	0.02	0.14	0	0.00	0	0	1	1	7.04	47.55	0.00

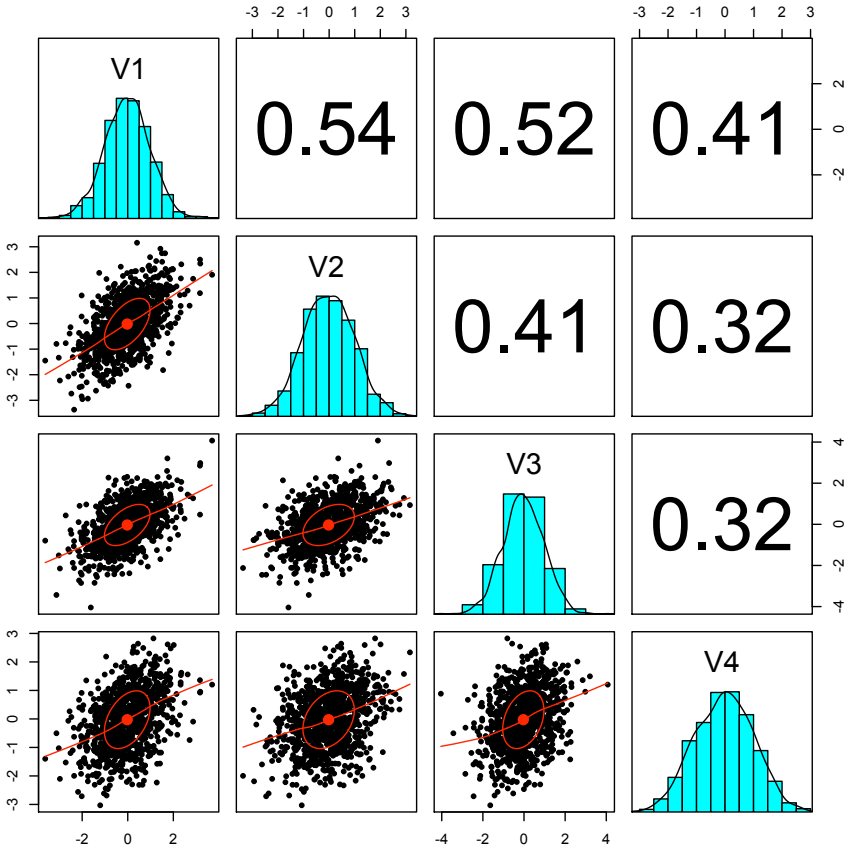
Phi coefficients



Tetrachorics as estimate of underlying correlation

```
> xc <- matrix(as.factor(x2),ncol=4)
> rc <- hetcor(xc)
> print(rc,digits=2)
Two-Step Estimates
Correlations/Type of Correlation:
     xc          X2          X3          X4
xc    1 Polychoric Polychoric Polychoric
X2 0.56           1 Polychoric Polychoric
X3 0.48        0.34           1 Polychoric
X4 0.45        0.33       -0.03           1
Standard Errors:
      xc    X2   X3
xc
X2 0.056
X3  0.07 0.086
X4   0.1  0.12 0.19
n = 1000
```

Compare with “true” data

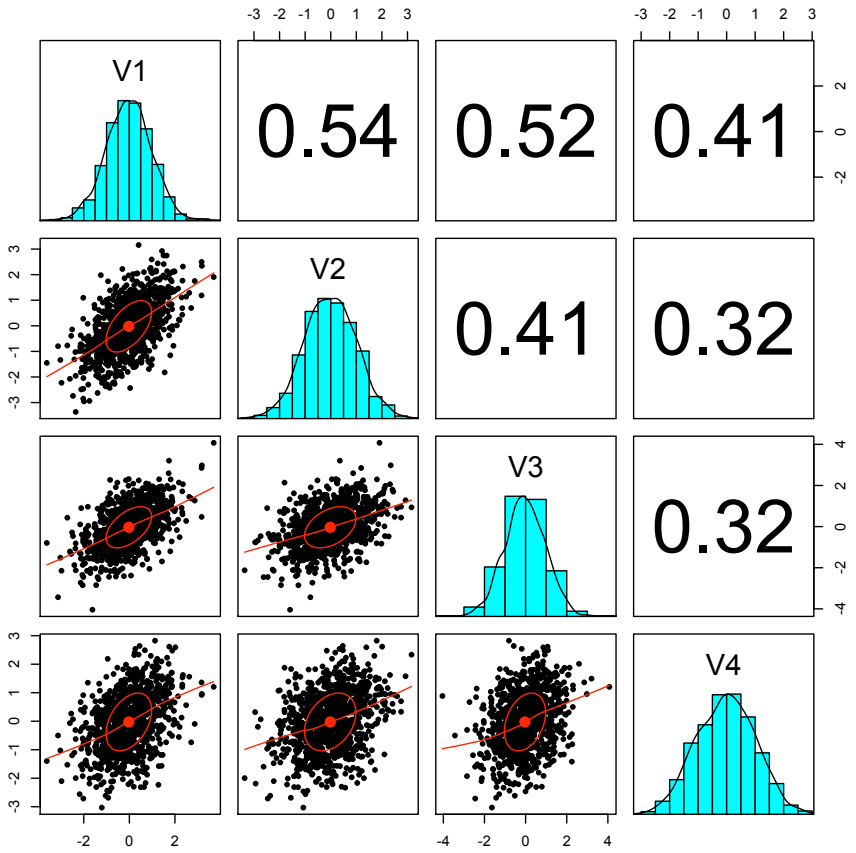



```
Correlations/Type of Correlation:
      xc          X2          X3          X
xc     1 Polychoric Polychoric Polychori
X2  0.56           1 Polychoric Polychori
X3  0.48        0.34           1 Polychori
X4  0.45        0.33       -0.03
```

If diagnoses are not as extreme, tetrachoric works better

```
set.seed(42)
x <- sim.congeneric(N=1000,short=FALSE)
cut <- rep(1,4)
x <- x$observed
x2 <- t((t(x)>cut)+0)

describe(x2)
   var    n mean   sd median trimmed mad min max range skew kurtosis   se
V1   1 1000 0.15 0.36      0    0.06   0   0   1     1 1.97     1.88 0.01
V2   2 1000 0.16 0.37      0    0.08   0   0   1     1 1.81     1.29 0.01
V3   3 1000 0.15 0.36      0    0.06   0   0   1     1 1.98     1.92 0.01
V4   4 1000 0.16 0.37      0    0.08   0   0   1     1 1.85     1.43 0.01
```

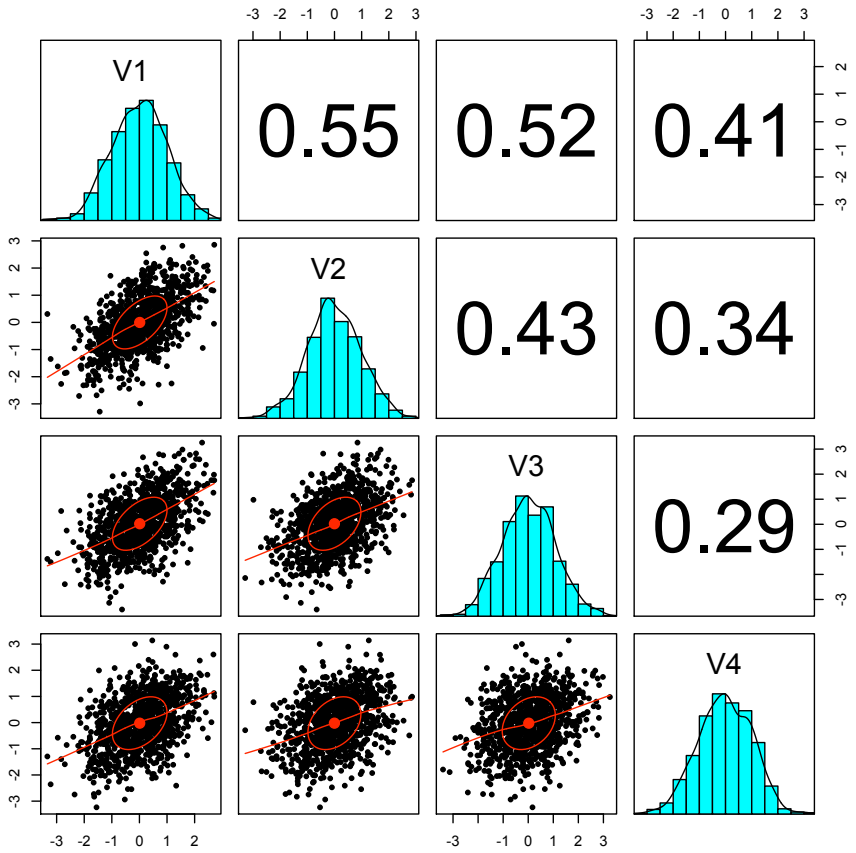
Underlying distribution



Convert to tetrachorics

```
> xc <- matrix(as.factor(xd),ncol=4)
> rtetra <- hetcor(xc)
> print(rtetra,digits=2)

Two-Step Estimates

Correlations/Type of Correlation:
     xc          X2          X3          X4
xc    1 Polychoric Polychoric Polychoric
X2 0.56           1 Polychoric Polychoric
X3 0.56        0.41           1 Polychoric
X4  0.1        0.19        0.15           1

Standard Errors:
      xc    X2   X3
xc
X2 0.054
X3 0.059 0.073
X4  0.15  0.15 0.17

n = 1000
```

Compare to generating data



	xc	X2	X3	X4
xc	1	Polychoric	Polychoric	Polychoric
X2	0.56	1	Polychoric	Polychoric
X3	0.56	0.41	1	Polychoric
X4	0.1	0.19	0.15	1

```
x2 <- sim.congeneric(N=1000,short=FALSE)$observed
> cut
[1] 1.0 1.2 1.5 2.0
xd <- t((t(x2) > cut)+0)
pairs.panels(x2)
```

Using tetrachorics

I. Convert comorbidities to correlations

II. find the structure of these correlations

III. Basically convert from categorical into a continuous model

Krueger and Markon, 2006