Problem Set 2

Carrie Kathlyn Townley Flores, Filipe Recch, Kaylee Tuggle Matheny, Klint Kanopka, Kritphong Mongkhonvanit, Paulina Julia Biernacki
EDUC 252L

February 4, 2018

1 Breaking the Classical Test Theory Model

1.1 Coin Flips

Coin flips should not be reliable data - they're random! To look at this a little more analytically:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} p_i(1-p_i)}{\sigma_X^2}\right)$$

The interesting thing to note here is that the probability of flipping heads is:

$$p_i = 0.5$$

And the variance on the sum of K coin flips will be:

$$\sigma_X = 0.25K$$

Substituting in the formula for Cronbach's Alpha:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}(0.5)(1-0.5)}{0.25K}\right)$$

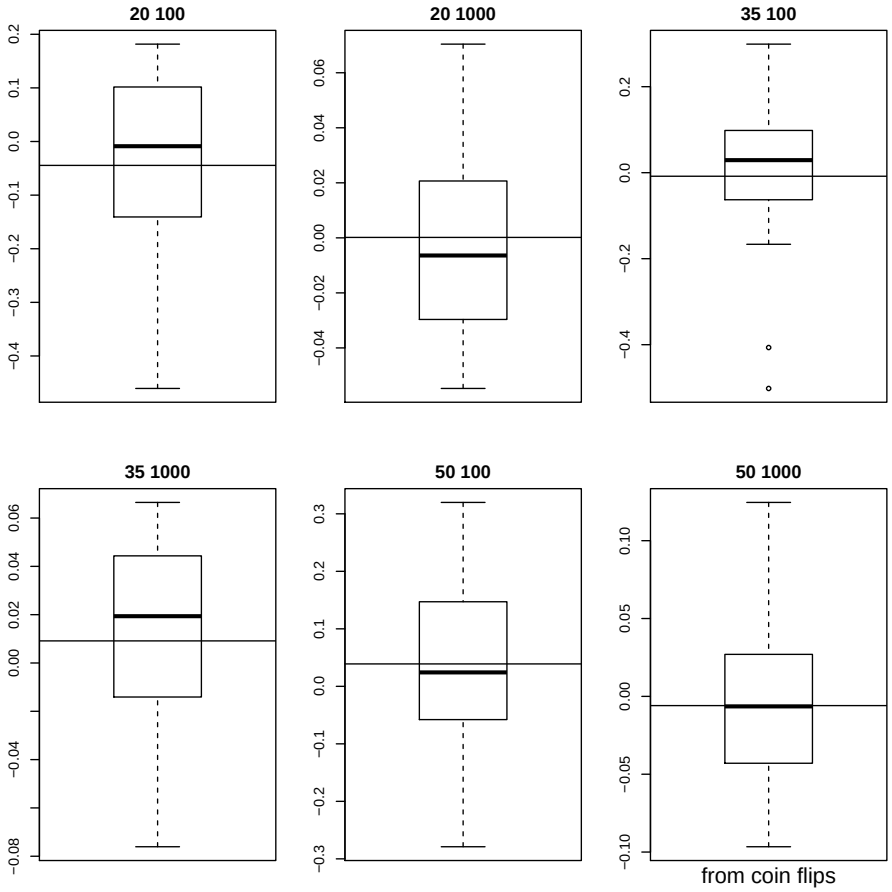
Cleaning up:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{0.25K}{0.25K}\right)$$

$$\alpha = \frac{K}{K-1}(1-1)$$

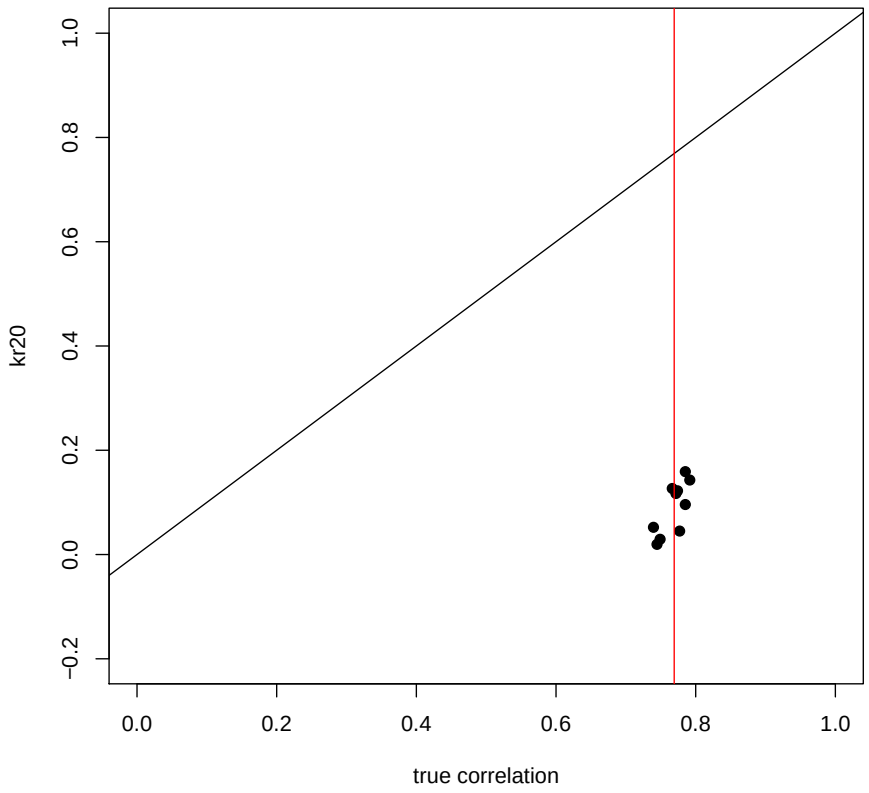
$$\alpha = 0$$

The expectation, then, is that α shoudl be zero for each situation.

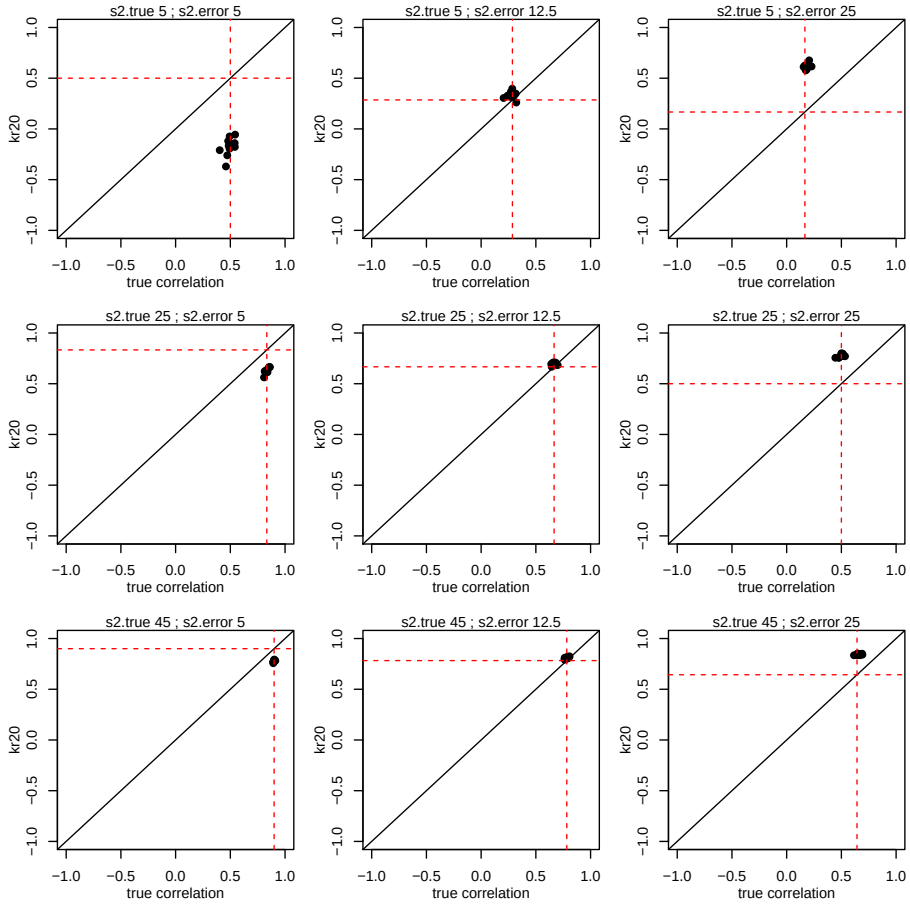


These α plots make sense - they are centered around zero, as predicted, and as the number of items increases, α is more tightly clustered around zero.

1.2 Simulating Item Response Data



The feature of the data generation mechanism that makes the α values super low is that items are getting marked correctly (essentially) at random! Even though the item responses generated correct p-values and test-level correlations, the data generation disregarded any internal structure you would expect. More clearly stated, respondents of similar ability levels did not have similar item response profiles.

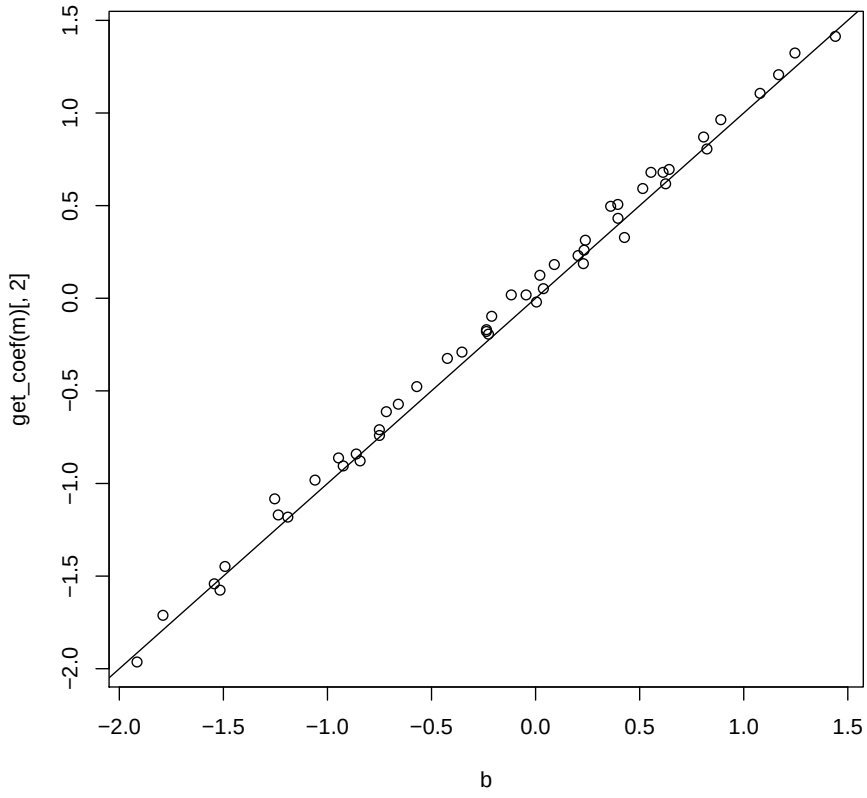


Looking at the resulting plots, it's clear that even with nonsensical item response data, the KR-20 estimate of reliability increases as a function of both true score variance and error variance. This *feels* very wrong. Increasing true score variance can be done by applying an instrument to a population it may not have been originally designed for. Increasing error variance can be done by adding more items or manipulating the quality of items. The challenge with feeling good about the CTT model is that KR-20 is both heavily valued and easily manipulated. The worst part is that some of the behaviors that would increase a KR-20 value could have negative impacts on the validity of the instrument.

2 Different Link Functions

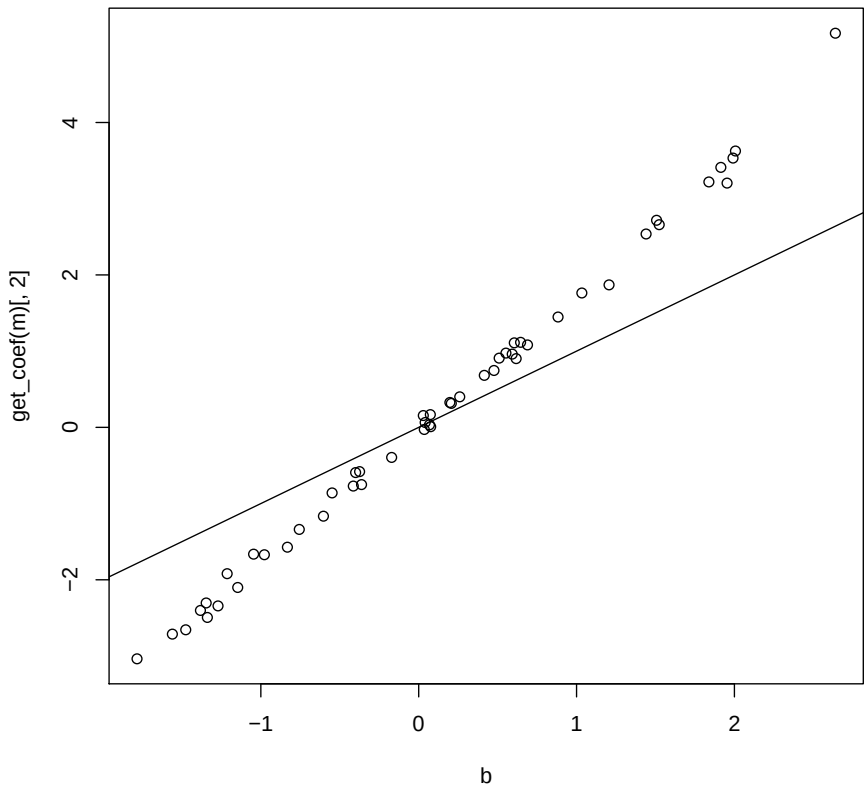
2.1 The Default

```
## Loading required package:  stats4
## Loading required package:  lattice
## Loading required package:  methods
```



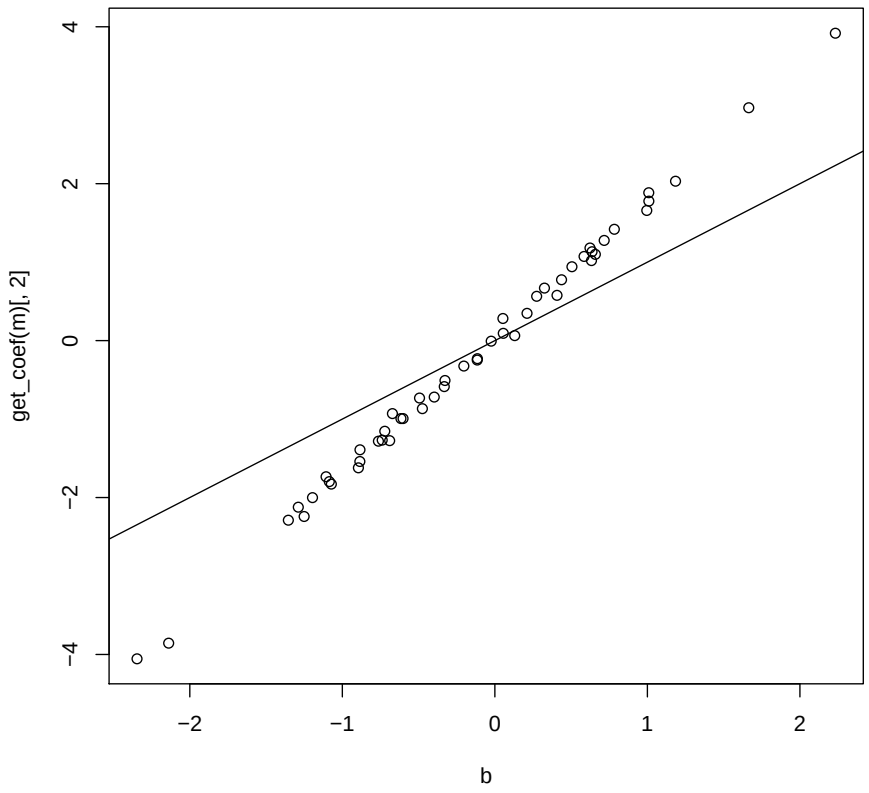
The default estimates item difficulties (or item easiness, specifically, because mirt) that are in line with the actual (specified) item difficulties.

2.2 The Normal



Using a normal link function, mirt predicts easy items are easier than they actually are and hard items are harder than they actually are. The farther an item is from zero, the larger the gap between estimated diffuculty and specified difficulty.

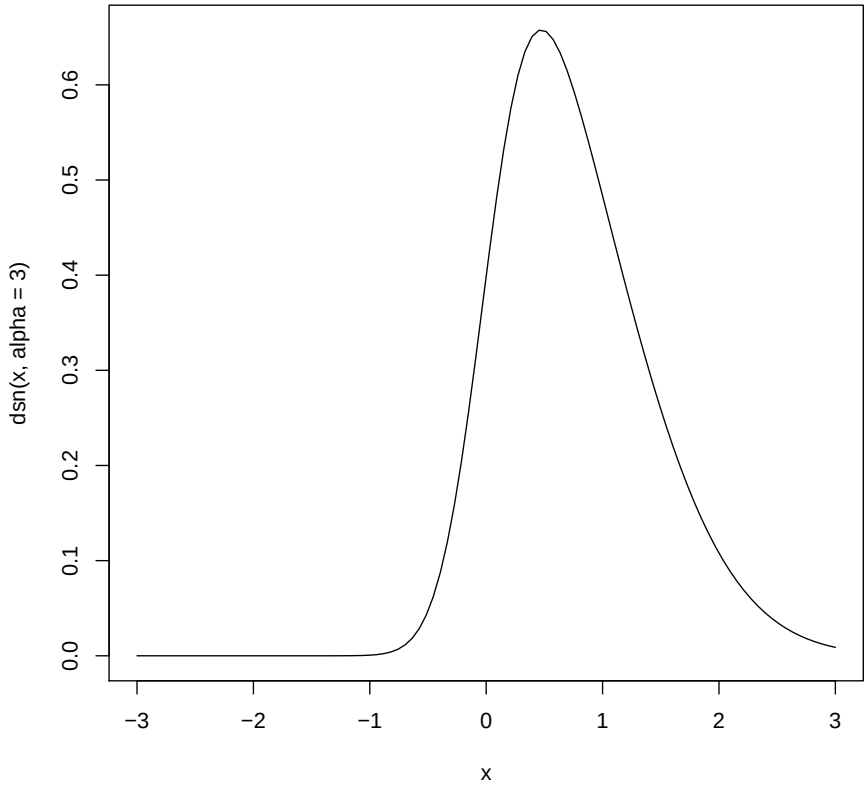
2.3 Heavy Tails



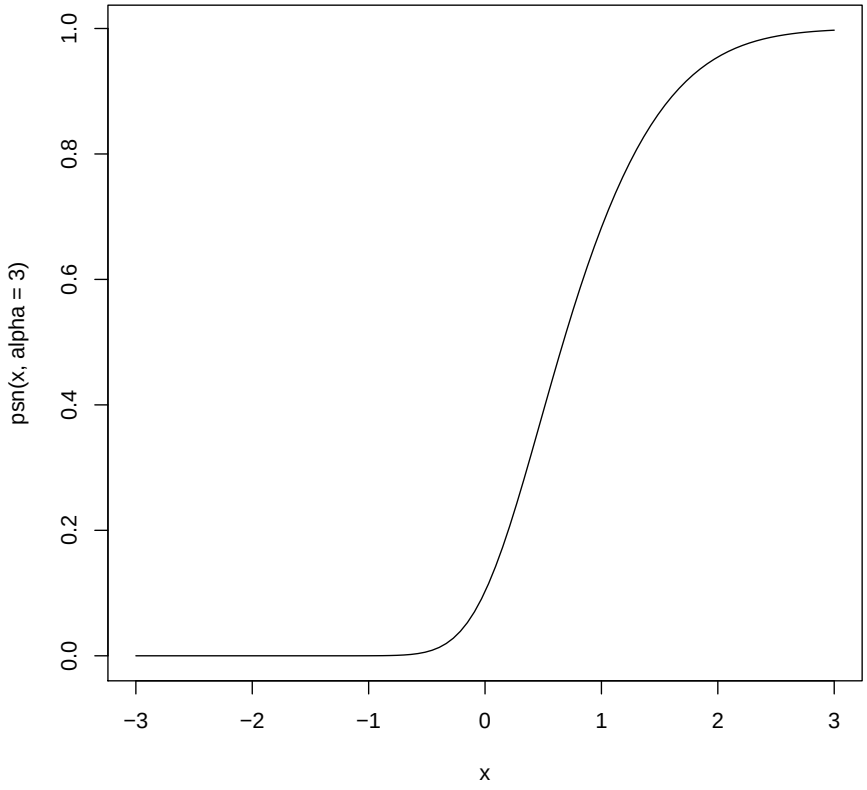
Using a link function with heavy tails, mirt does the same as above.

2.4 Skewed

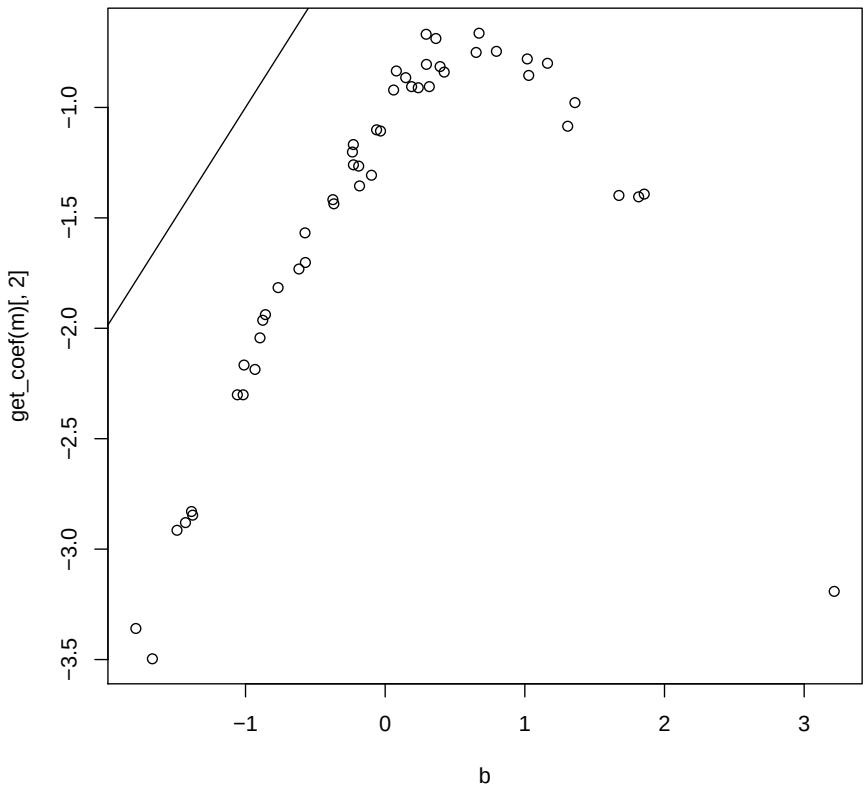
```
##
## Attaching package:  'sn'
## The following object is masked from 'package:stats':
##
##    sd
```



53



54

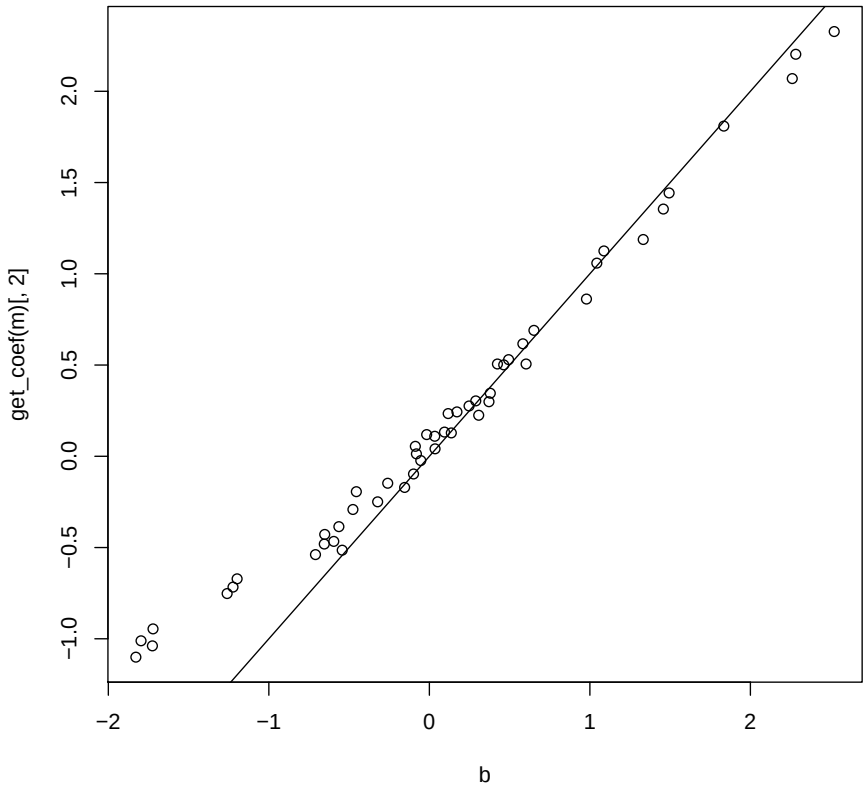


When using a skewed distribution as the link function, every item is estimated as being harder than it actually is, but where this model really tanks is for extremely easy items. It finds those to be significantly harder than they actually are and the difference between actual and estimated difficulty for easy items diverges wildly. The skew on the link function is set up in a way that there is essentially no impact of increased θ on the respondent's probability of getting the item correct until θ gets to a value of -1. Remember that the input to the model is of the form $\theta_i + b_j$. If trait values less than -1 don't provide any increase in probability of scoring correctly on an item, for items with easiness above 1, the model will view those as significantly harder than they actually are. If you look at the $b_estimated$ vs. b curve, you can see that at $b = 1$, the estimated easiness starts to take a turn away from actual easiness.

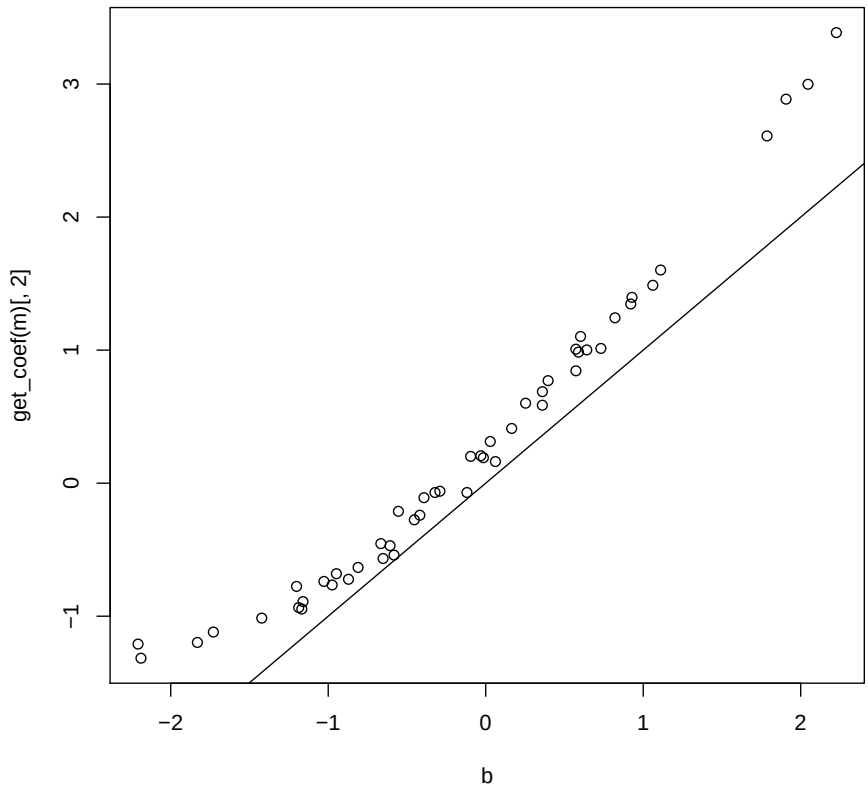
3 Simulated IRT Data

3.1 The Default

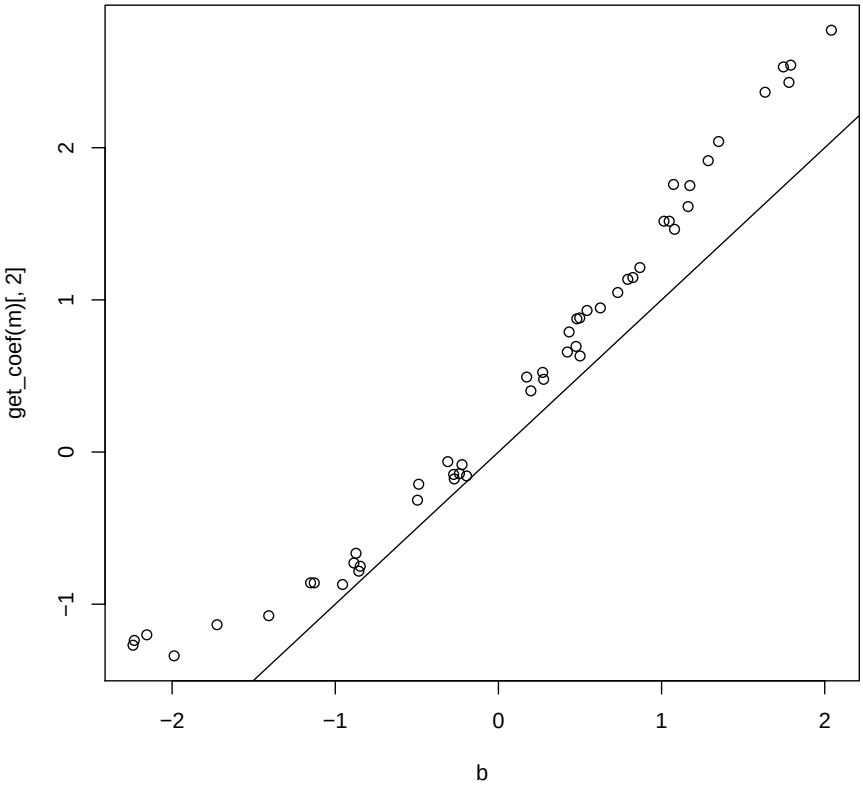
```
##
Iteration: 1, Log-Lik: -29832.251, Max-Change: 0.24536
Iteration: 2, Log-Lik: -29802.480, Max-Change: 0.06158
Iteration: 3, Log-Lik: -29799.897, Max-Change: 0.01843
Iteration: 4, Log-Lik: -29799.646, Max-Change: 0.00557
Iteration: 5, Log-Lik: -29799.605, Max-Change: 0.00221
Iteration: 6, Log-Lik: -29799.602, Max-Change: 0.00070
Iteration: 7, Log-Lik: -29799.601, Max-Change: 0.00027
Iteration: 8, Log-Lik: -29799.601, Max-Change: 0.00022
Iteration: 9, Log-Lik: -29799.601, Max-Change: 0.00006
```



3.2 The Normal

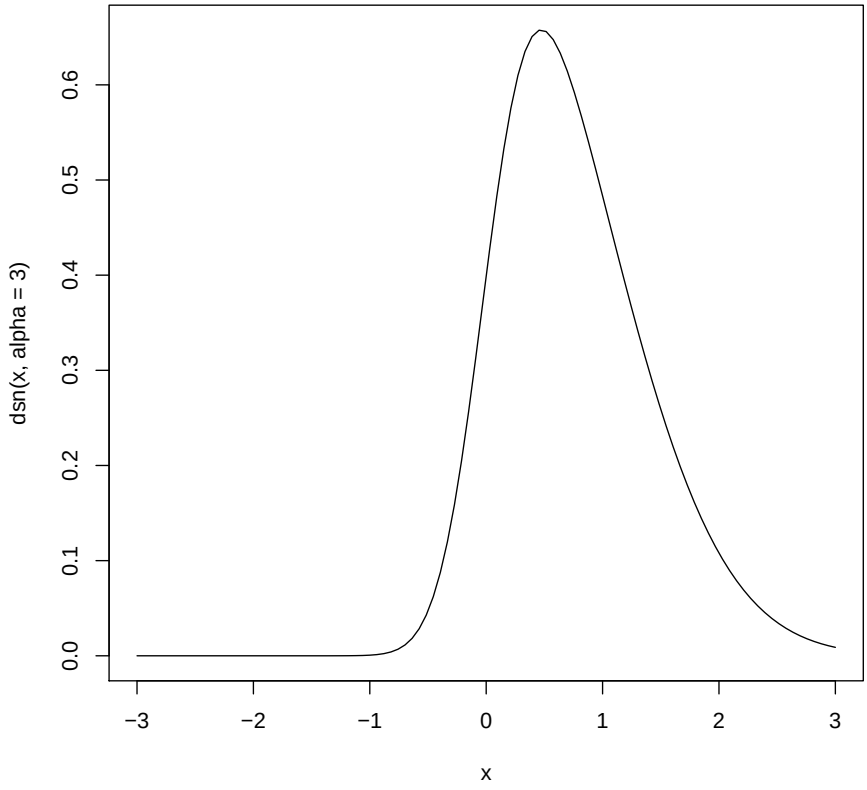


3.3 Heavy Tails

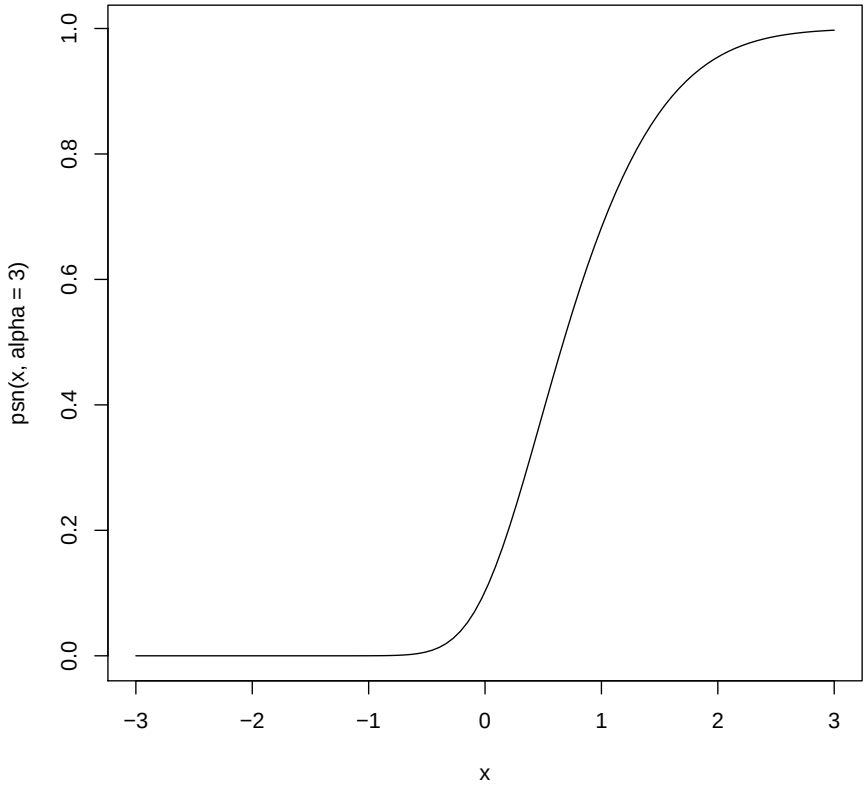


74

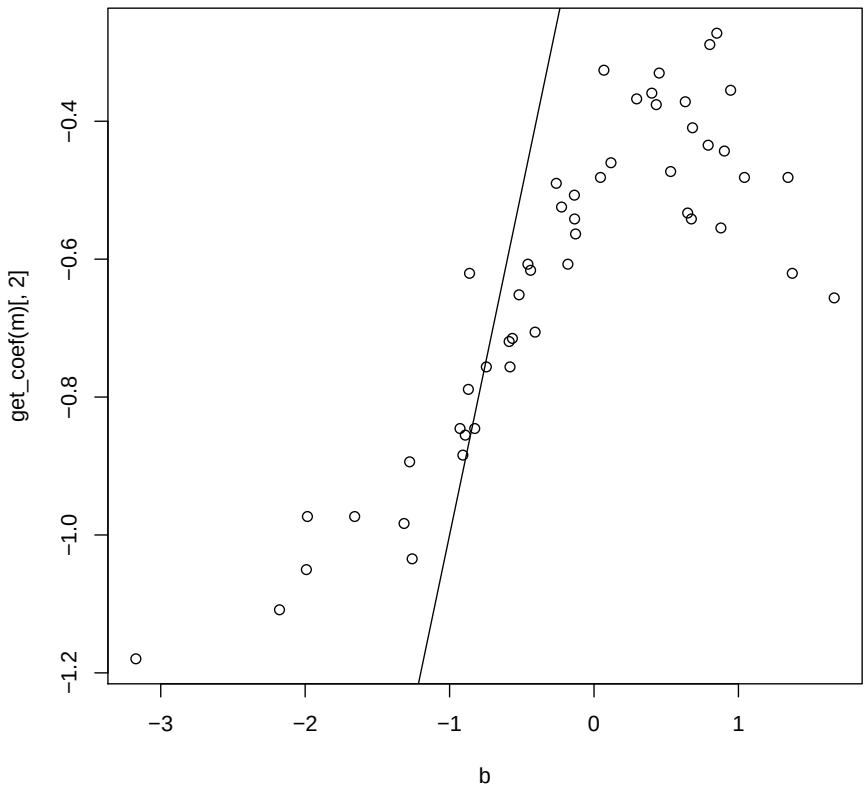
75 3.4 Skewed



76



77



78