Sorites Model Fits

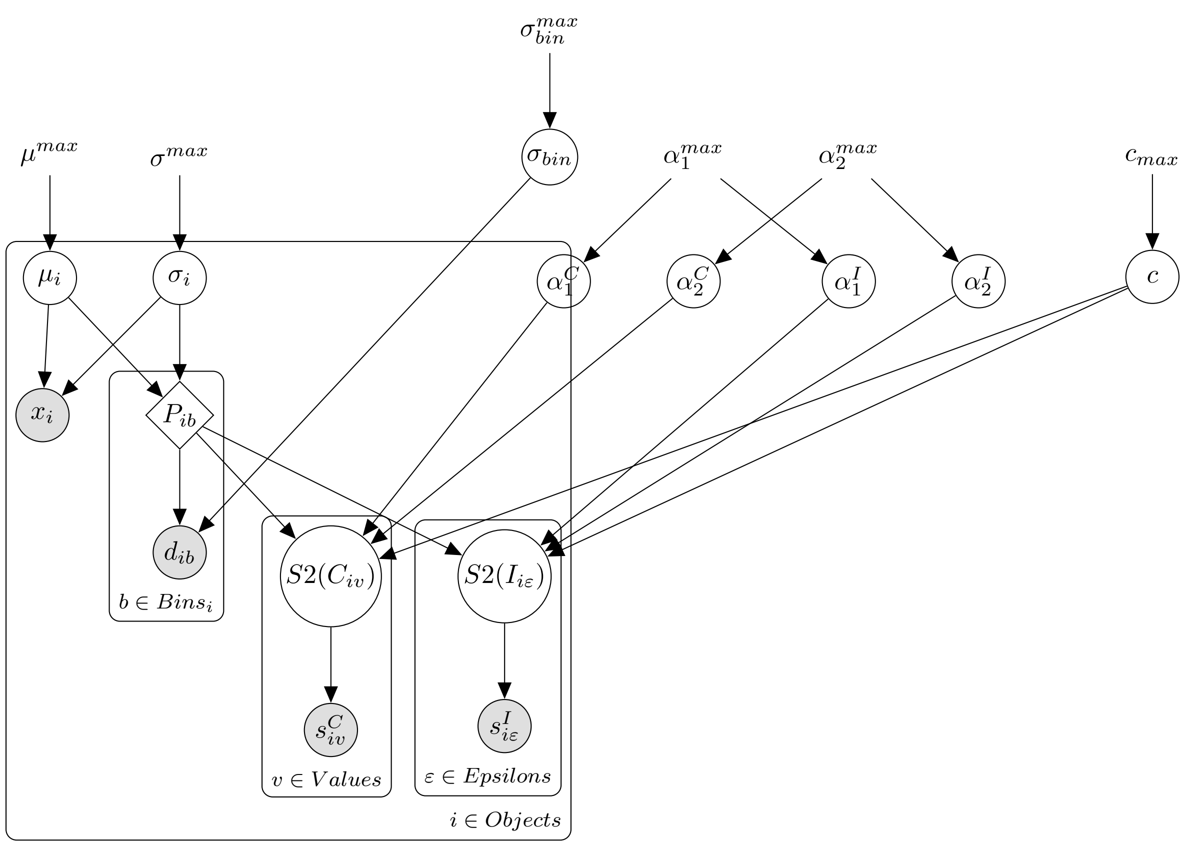
Erin Bennett

Model

Definitions

- w_i := width of histogram bins for item i
- x_i := sample in give a number trial
- P_{ib} := true probability of bin b for item i
- d_{ib} := slider rating for bin b for item i
- $S2(I_{i\varepsilon})$:= RSA S2(L1(expensive) + ε) for item i
- $S2(C_{iv})$:= RSA S2(expensive) for item i
- $s^I_{i\varepsilon}$:= binarization of likert rating for inductive premise for item i and epsilon ε
- s^C_{iv} := binarization of likert rating for concrete premise for item i and value v
- α^I_2 := speaker rationality for S1 for inductive premise

Diagram



Distributions/Functions/Values:

Experiment design parameters:

- $Objects$
- $Bins$
- $Epsilons$
- $Values$

Assumed model parameters:

- $\mu^{max} = 20$
- $\sigma^{max} = 5$
- $\sigma_{binned\ hist} = ??$
- $\alpha_1^{max} = 20$
- $\alpha_2^{max} = 5$
- $\sigma_{bin}^{max} = 5$

Inferred Latent variables:

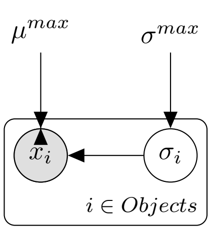
- $\mu_i \sim \mathcal{U}\{0, \mu^{max}\}$
- $\sigma_i \sim \mathcal{U}\{0, \sigma^{max}\}$
- $\alpha_1^I \sim \mathcal{U}\{0, \alpha_1^{max}\}$
- $\alpha_2^I \sim \mathcal{U}\{0, \alpha_2^{max}\}$
- $\alpha_1^C \sim \mathcal{U}\{0, \alpha_1^{max}\}$
- $\alpha_2^C \sim \mathcal{U}\{0, \alpha_2^{max}\}$
- $\sigma_{bin} \sim \mathcal{U}\{0, \sigma_{bin}^{max}\}$
- $P_{ib} = \int_{LB_{ib}}^{UB_{ib}} \varphi(\ln(t)|\mu_i, \sigma_i)dt$

Observations from experimental data:

- $logit(d_{ib}) \sim \mathcal{N}(logit(p_{ib}), \sigma_{bin})$
- $\ln(x_i) \sim \mathcal{N}(\mu_i, \sigma_i)$
- s_{iv}^C
- $s_{i\varepsilon}^I$

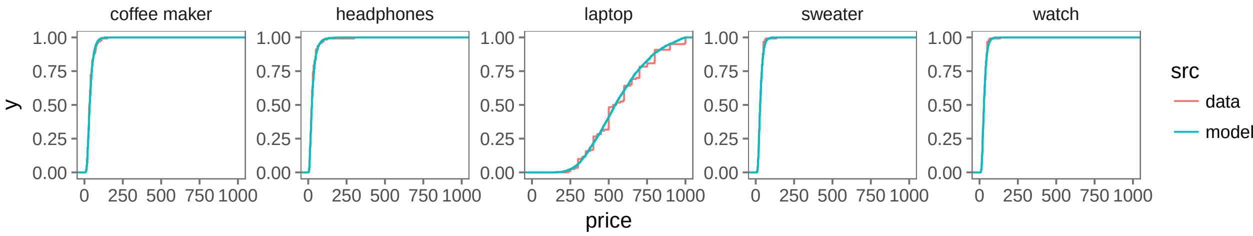
Model fit

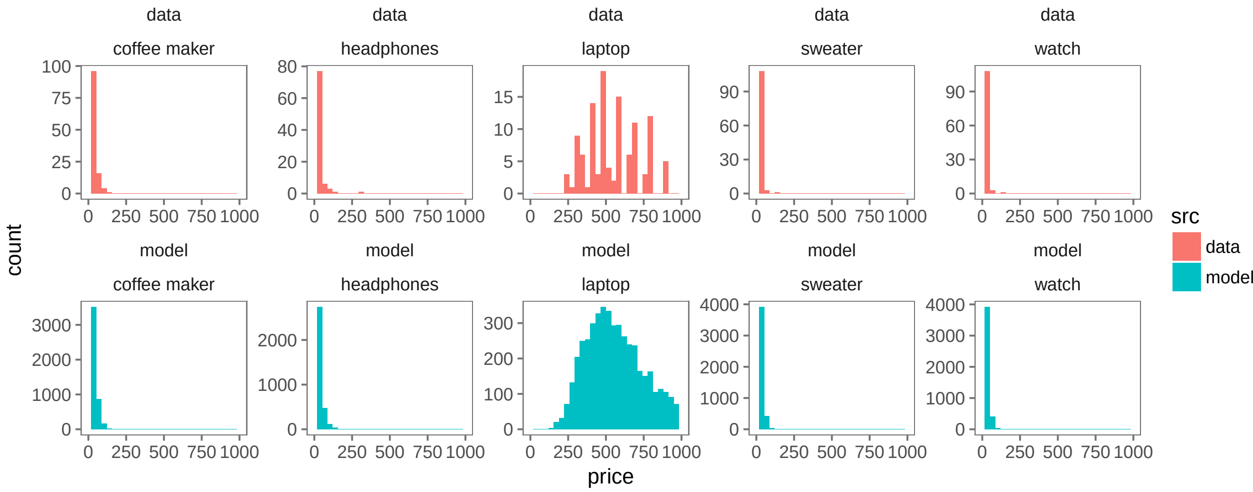
Give a Number ~ Log Normal

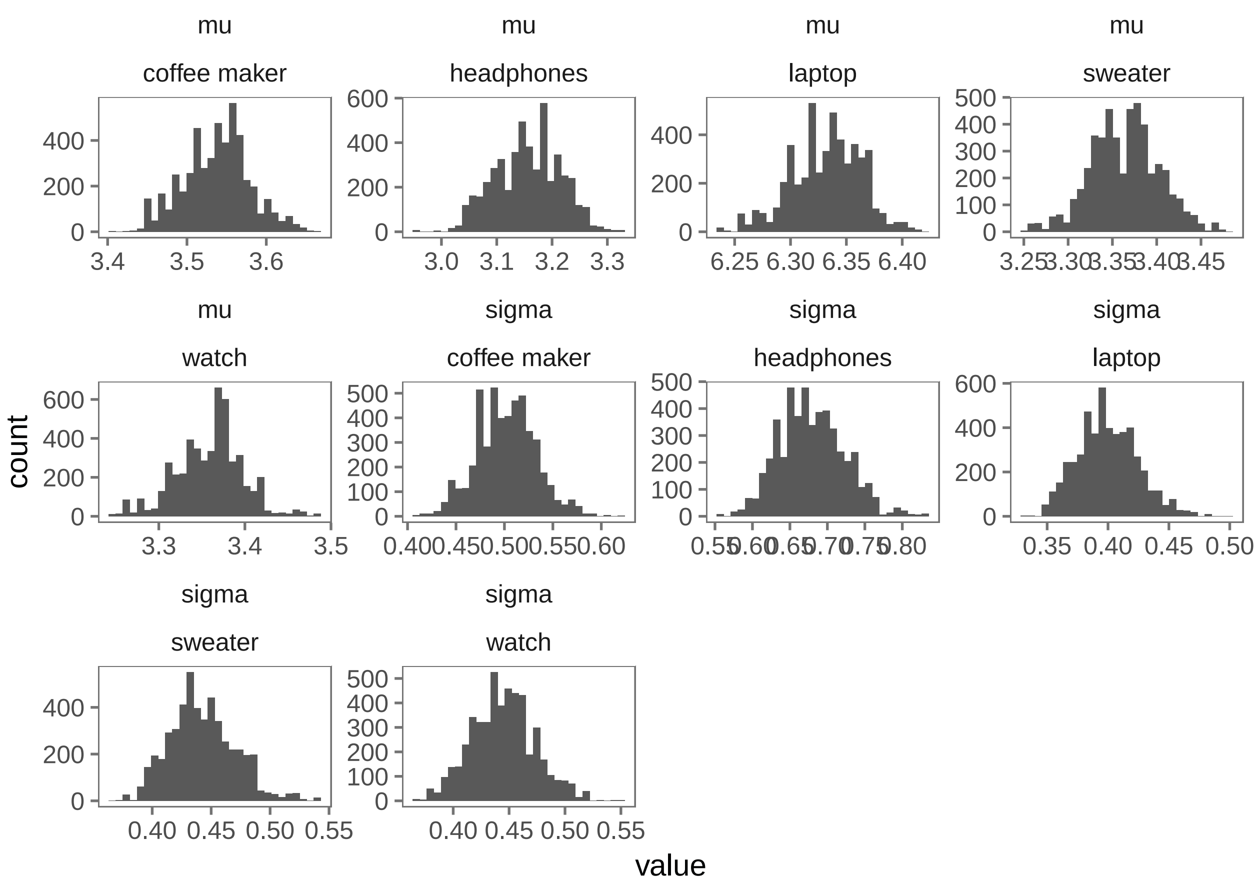


Model fit for Give a Number as log normal using Incremental MH.

```
iterations = 5000
burn = iterations/2
lag = 10
```


coffee maker
headphones
laptop
sweater
watch
y
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
price
src
data
model


data
coffee maker
headphones
laptop
sweater
watch
model
count
price
src
data
model


mu
coffee maker
headphones
laptop
sweater
watch
sigma
count
value

Concrete Premise ~ lifted L1

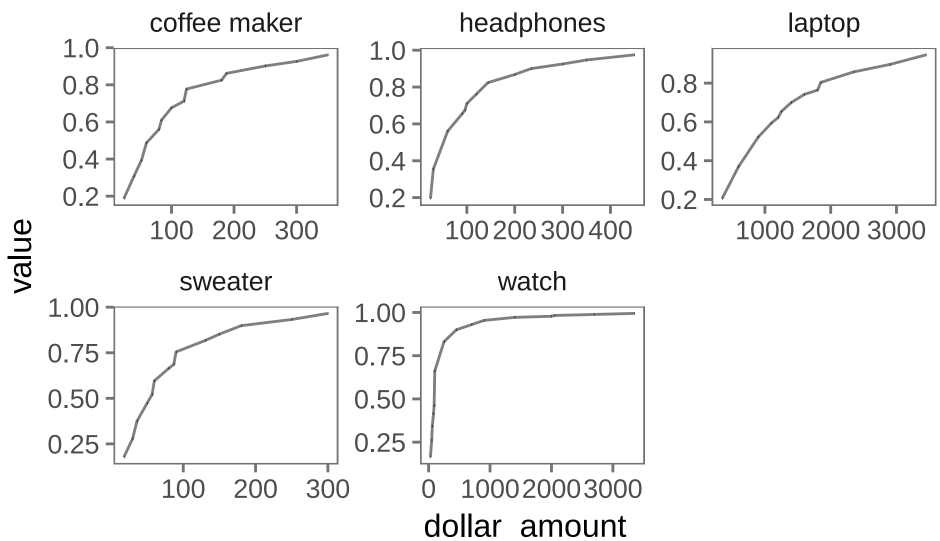
I'm discretizing into bins independently and when price x and threshold θ are in the same bin, I'm using a `flip()` to decide whether $x \leq \theta$ or $theta < x$. This isn't quite right, since the shape of the x distribution on the interval affects the proportion of the time that $theta < x$. But it's an approximation that works pretty well and shouldn't mess anything up too much. I spent a lot of time thinking about the true joint distribution in `discretization.html` but ultimately this version, with `flip`, is the one that generated these results.

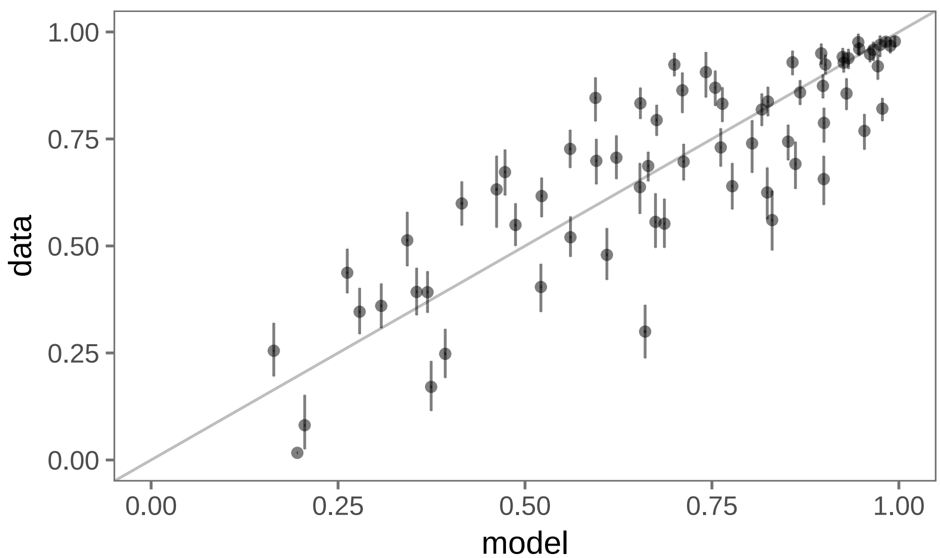
```
iterations = 1000
burn = iterations/2
lag = 10
```

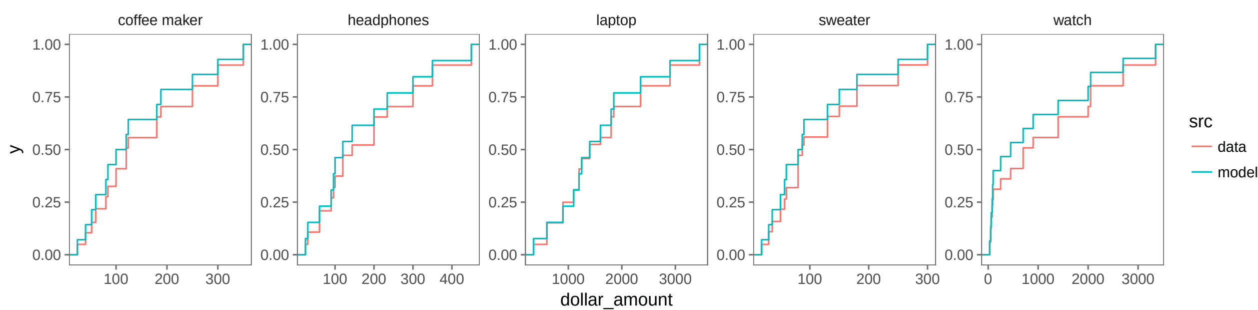
```
## # A tibble: 13 x 2
##    dollar_amount value
##            <dbl> <dbl>
##  1          350. 0.205
##  2          600. 0.370
##  3          900. 0.522
##  4         1100. 0.594
##  5         1200. 0.622
##  6         1250. 0.654
##  7         1400. 0.700
##  8         1600. 0.742
##  9         1800. 0.764
## 10         1850. 0.804
## 11         2350. 0.858
## 12         2900. 0.896
## 13         3450. 0.946
```

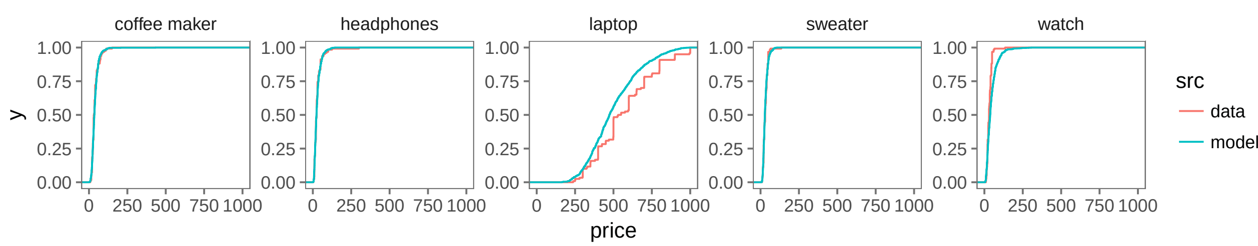
```
## # A tibble: 13 x 2
##    dollar_amount value
##            <dbl> <dbl>
##  1          350. 0.205
##  2          600. 0.370
##  3          900. 0.522
##  4         1100. 0.594
##  5         1200. 0.622
##  6         1250. 0.654
##  7         1400. 0.700
##  8         1600. 0.742
##  9         1800. 0.764
## 10         1850. 0.804
## 11         2350. 0.858
## 12         2900. 0.896
## 13         3450. 0.946
```

memoized `listener1` and `listener1_score`, 500 iterations, lag 10:


coffee maker
headphones
laptop
sweater
watch
value
dollar_amount


data
model


coffee maker
headphones
laptop
sweater
watch
y
dollar_amount
src
data
model


coffee maker
headphones
laptop
sweater
watch
y
price
src
data
model