Motivation

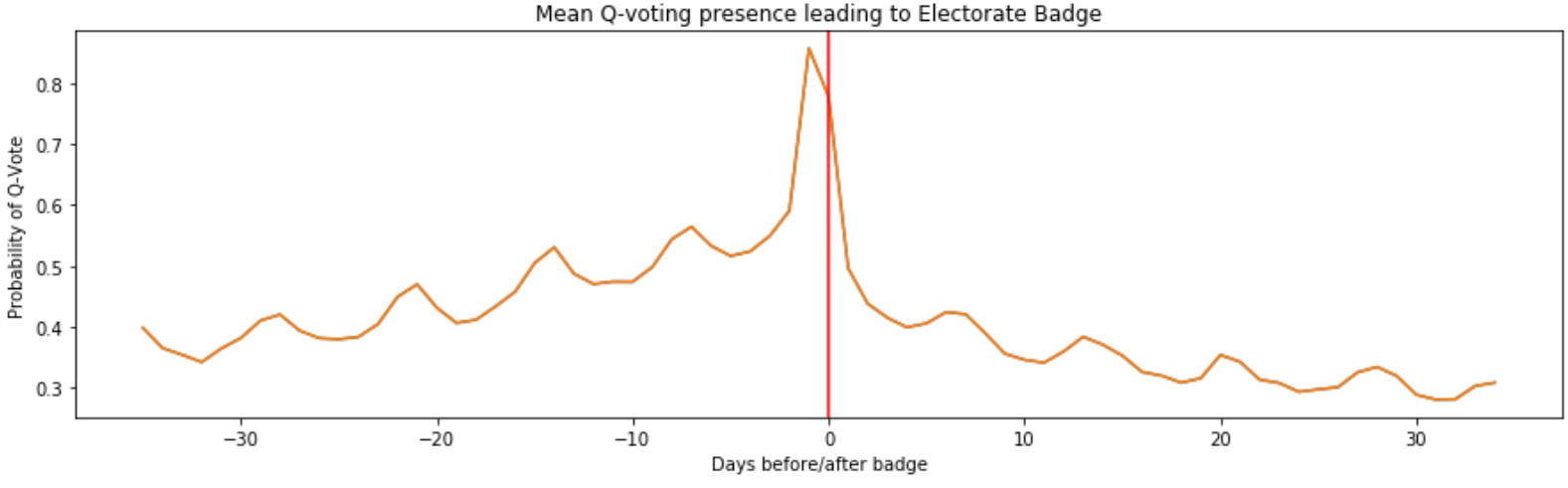
We extend the theoretical model from **Anderson et al. (2013), Steering User Behavior with Badges** to a model that is grounded in the users' behaviour on Stack Overflow. From the model, we gain insights about how people act in the presence of a badge and how they repond to badge incentives.

Original model:

$$U(x_a) = \sum_{b \in B} I_b(a) V_b + \theta \sum_{i=1}^{n+1} x_a^i \cdot U(x_{a+e_i}) - g(x_a, \mathbf{p})$$

The probability of acting depends on:

1. The user specific base distribution ($\mathbf{p}$).
2. The value of a badge (not specified if this is user specific or assumed external to the user) (V_b).
3. The cost that a user pays for deviating from the base distribution (g).
4. (some discount parameter θ)

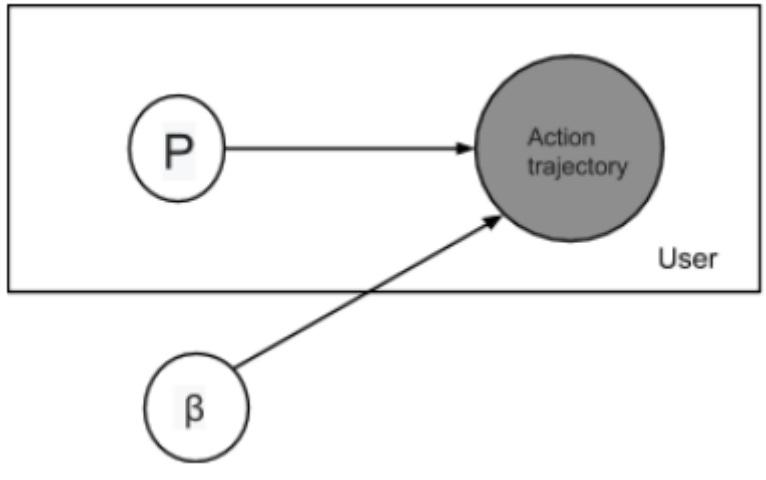

Mean Q-voting presence leading to Electorate Badge
Probability of Q-Vote
0.8
0.7
0.6
0.5
0.4
0.3
−30
−20
−10
0
10
20
30
Days before/after badge

Express as a graphical model

- A user will only deviate if the badge value outweighs the cost of changing his/her default behaviour.
- It follows that (2) and (3) can be understood as a deviation from the user's base distribution as the user approaches the badge boundary. This deviation is non-negative before the badge is achieved and non-positive after the badge has been achieved.
- **A simplifying assumption:** the "badge deviation" (called kernel from now on) is a function that depends on the user's proximity to the badge and is a shared response between all users. This assumption can be relaxed at a later stage.

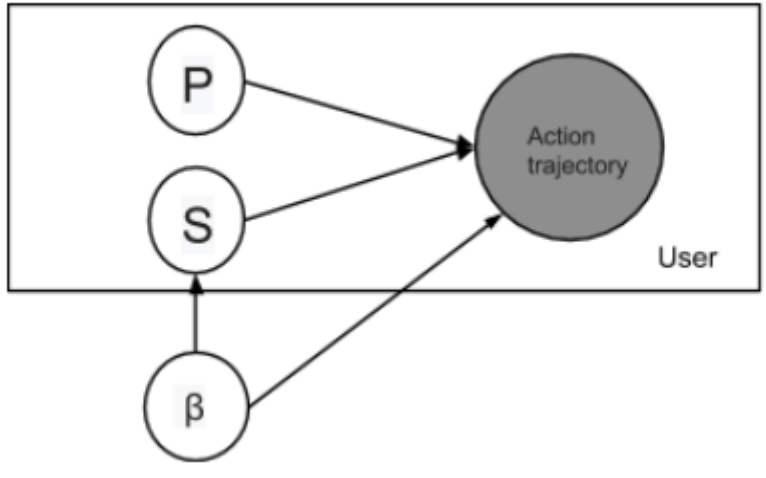
Express as a graphical model

We obtain the following graphical model for a user's interaction behaviour (p is the base distribution local to a user and β is the badge kernel:



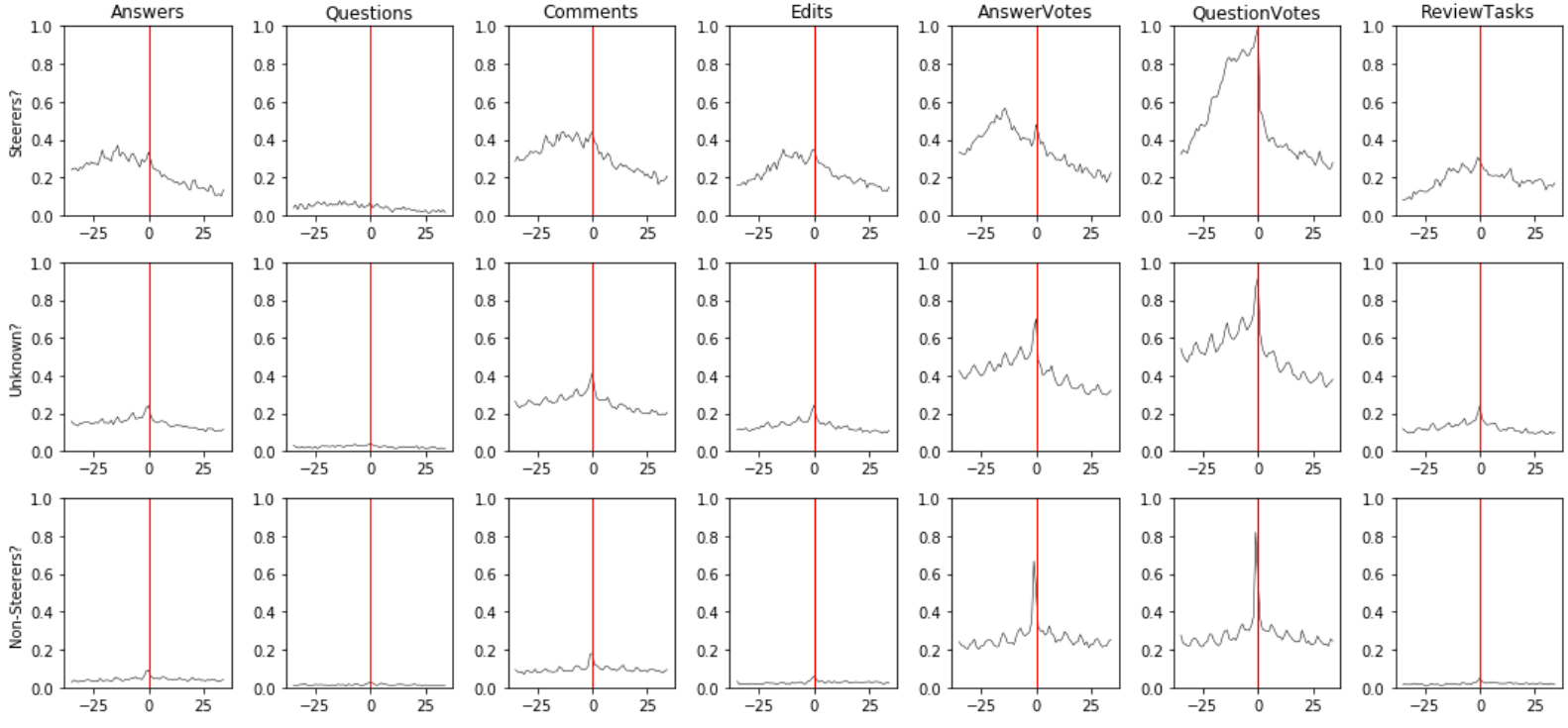
Address Weaknesses of baseline model:

There is evidence that users do not respond uniformly to the badge. **Solution**: add a user-specific parameter that controls the strength of the effect of the badge response (shown below). $S \in (0, 1)$ is the parameter that controls the strenth of a user's adherence to the kernel. P and S are both local latent variables, specific to each user.



Note: It is possible that users do not respond with the same functional form to a badge (not addressed yet).

Brief motivation for strength parameter (S):


Answers
Questions
Comments
Edits
AnswerVotes
QuestionVotes
ReviewTasks
Steerers?
Unknown?
Non-Steerers?
1.0
0.8
0.6
0.4
0.2
0.0
−25
0
25

Notes:

- Different populations have different responses to the badge in the proximity of the badge.
- The population that starts with the lowest count of actions (in the 5 week period) has the biggest deviance from 0 in their response

Method

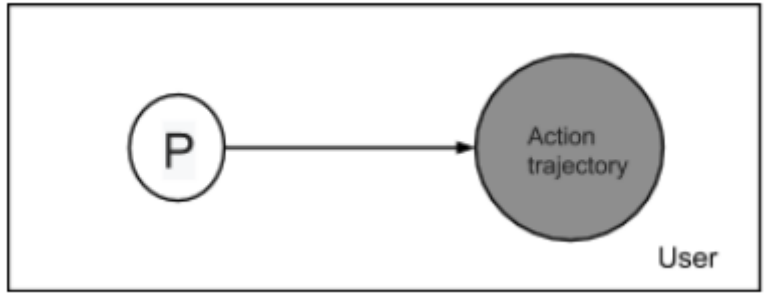
Compare and contrast the models, investigate the functional form of the steering kernel and relax modeling assumptions.

Baseline (model 1)

Model assumptions:

(1) Every user has their own "base distribution".

(2) Base distribution does not change between weeks of interaction.




```
Average Test loss: 40.7442
```

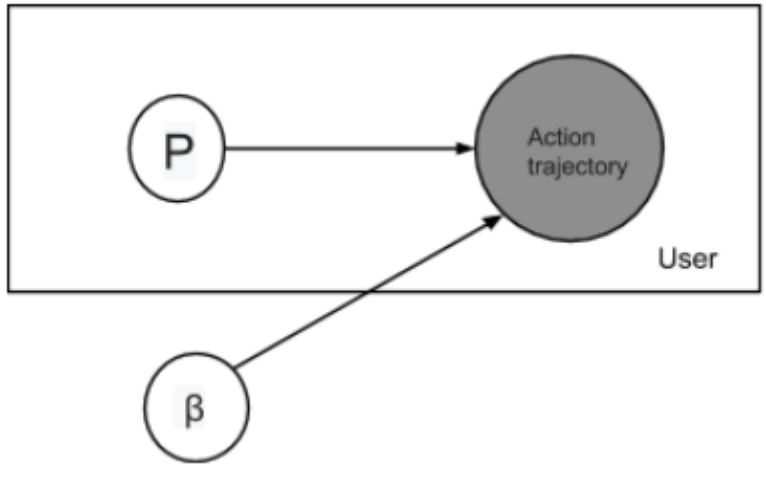
Linear parameter (model 2)

Model assumptions:

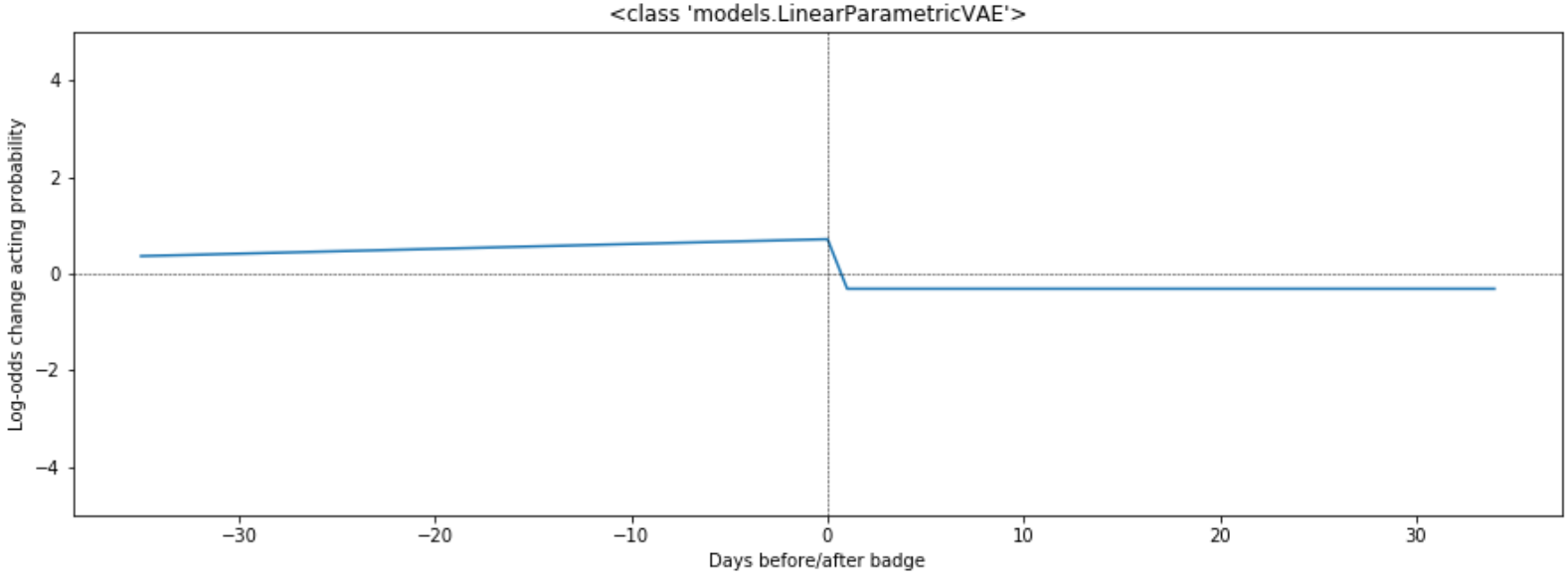
(1) Every user has their own "base distribution".

(2) Base distribution does not change between weeks of interaction.

(3) All users experience a linear strictly positive change in probability of acting as they approach the badge boundary. Thereafter they experience a non-positive change in probability in acting.



Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2c957e48>

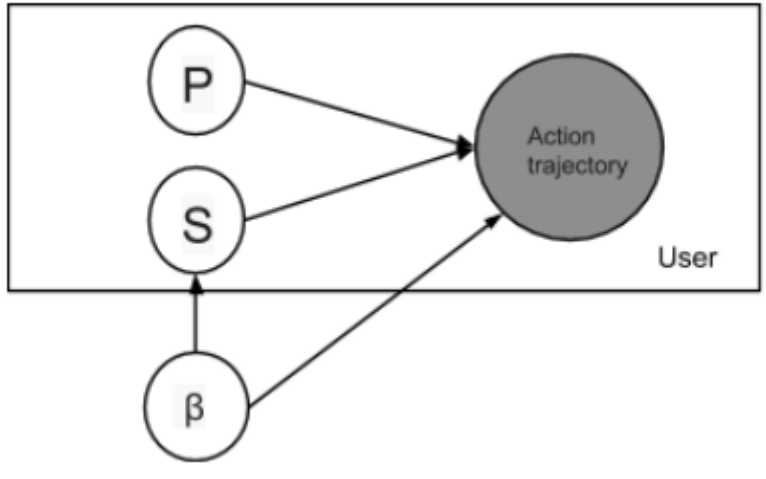


Average Test loss: 40.1143

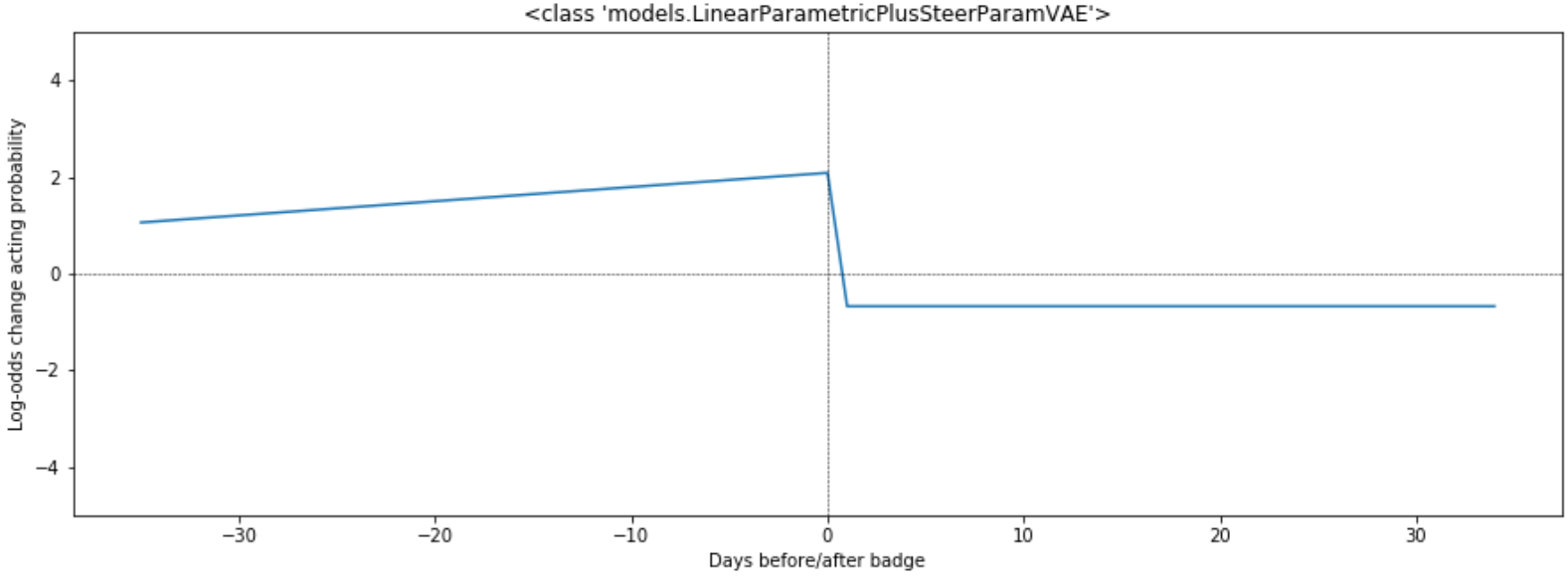
Personalised Linear (model 3)

Model assumptions - same as Linear but:

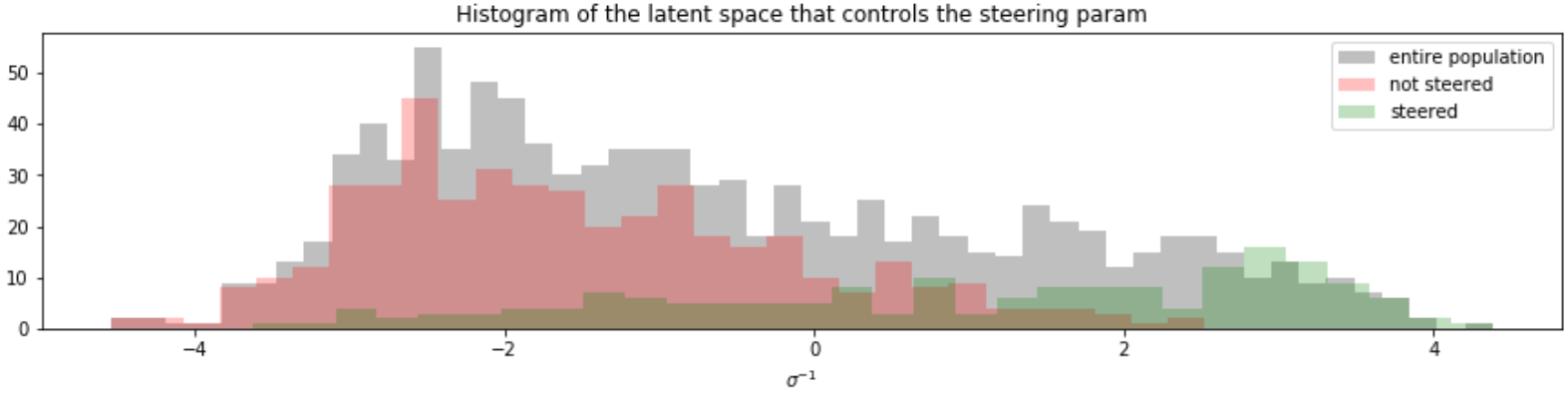
1) A user-specific steering parameter $\in (0, 1)$ that controls the effect of the deviance from normal acting.

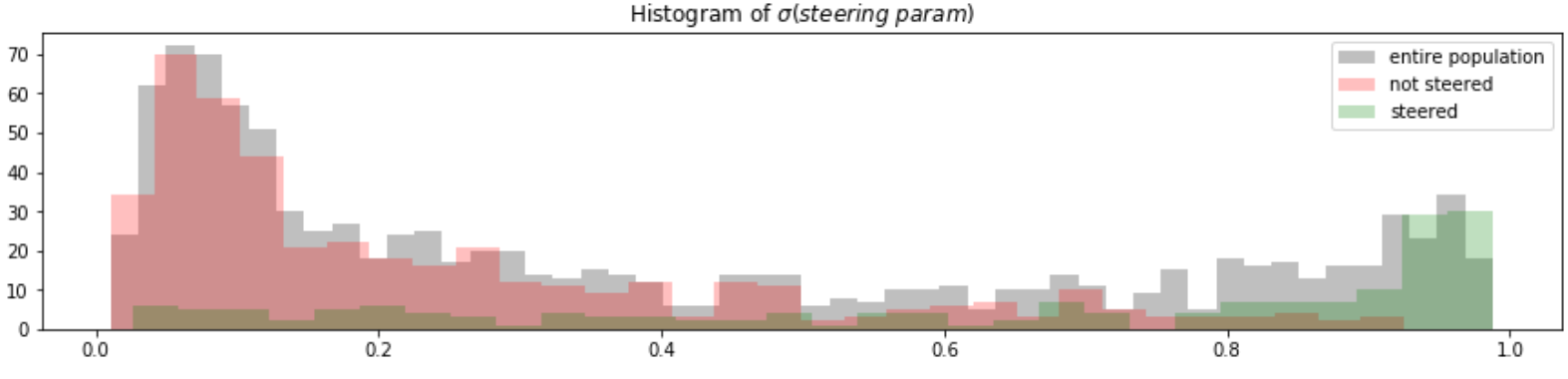


Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2c68ddd8>



Average Test loss: 39.2619


Histogram of the latent space that controls the steering param
entire population
not steered
steered
50
40
30
20
10
0
−4
−2
0
2
4
σ^{-1}


Histogram of σ(*steering param*)
entire population
not steered
steered
70
60
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0

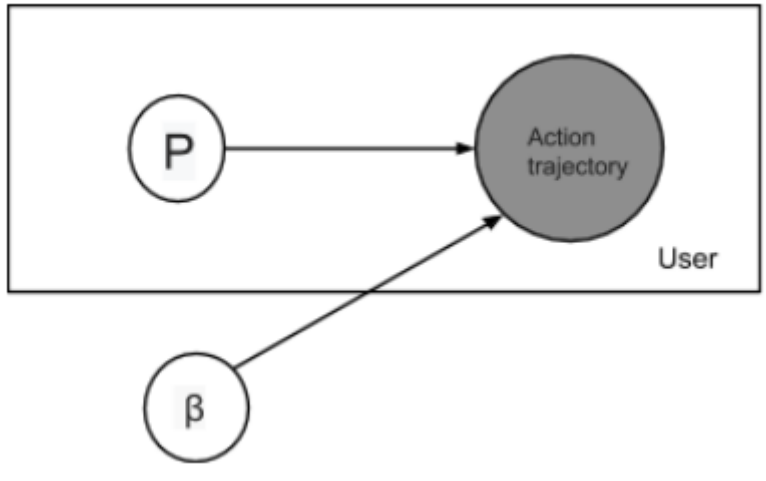
Fully parameterised (model 4)

Model assumptions:

(1) Every user has their own "base distribution".

(2) Base distribution does not change between weeks of interaction.

(3) All users experience a non-negative change in probability of acting before the badge is achieved and a non-positive change in probability of acting before the badge is achieved.



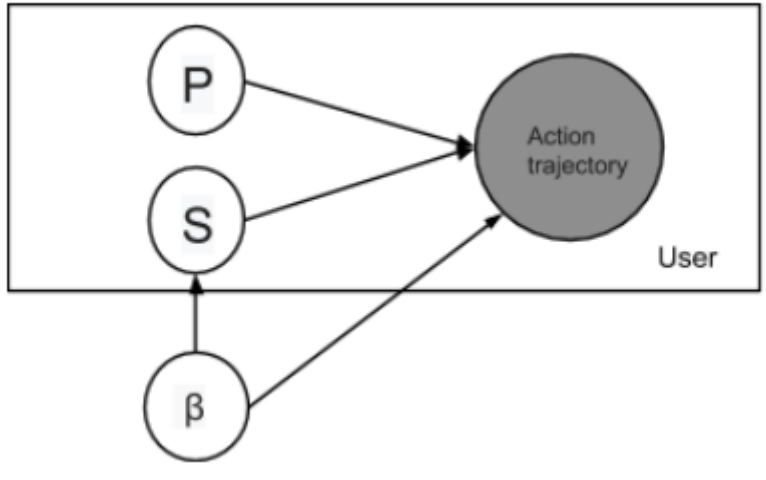
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2d366ba8>

Average Test loss: 39.0297

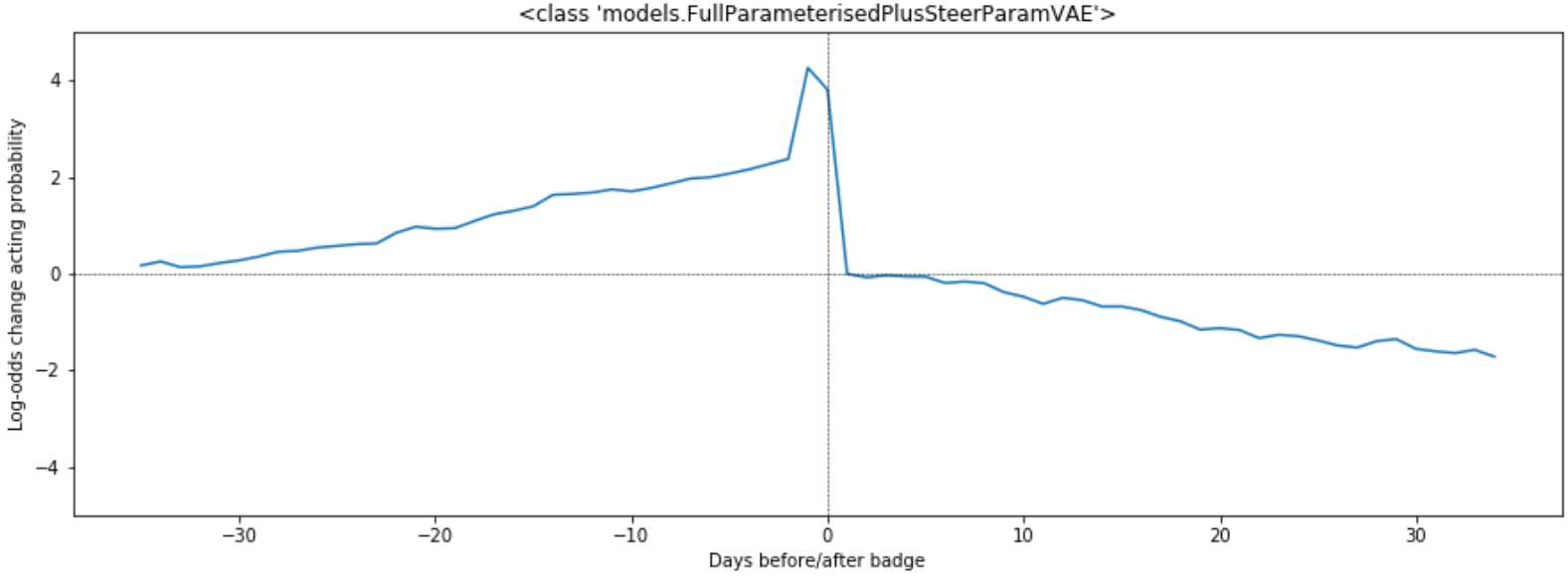
Fully parameterised, personal param (model 5)

Model assumptions - same as full parameterised but:

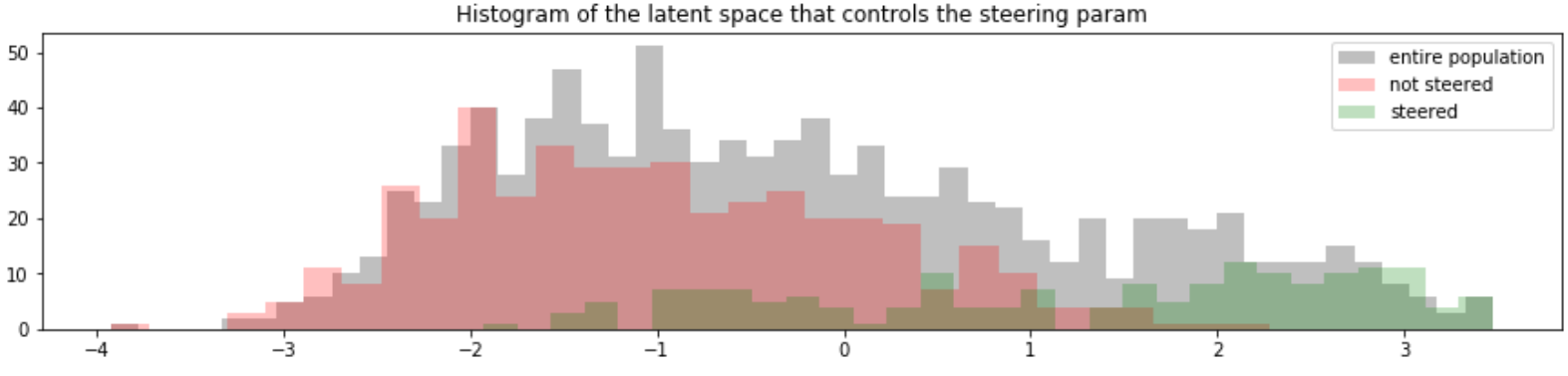
1) A user-specific steering parameter $\in (0, 1)$ that controls the effect of the deviance from normal acting.



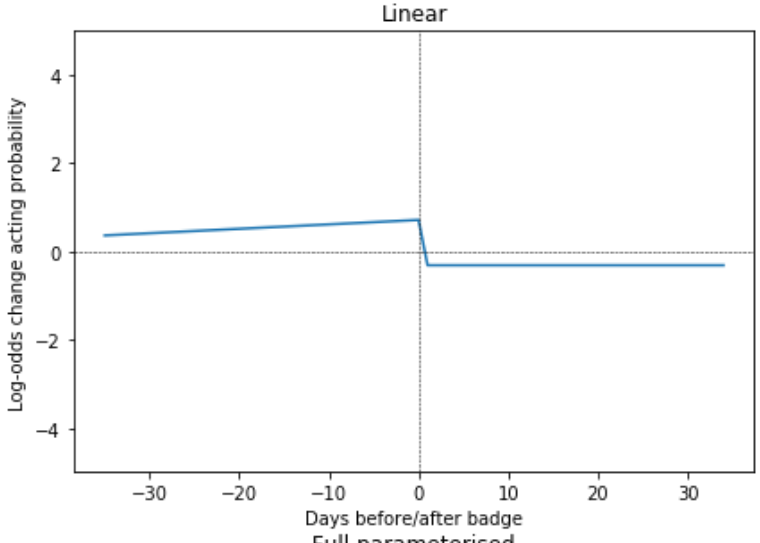
Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x1a31096780>

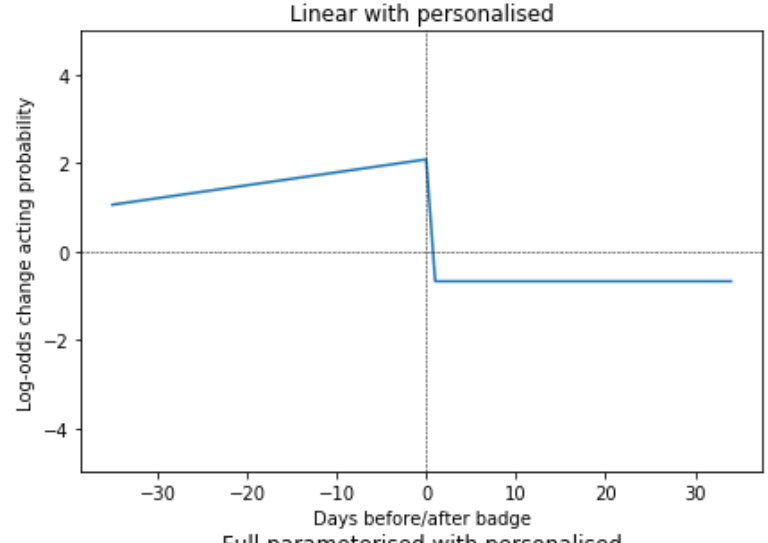


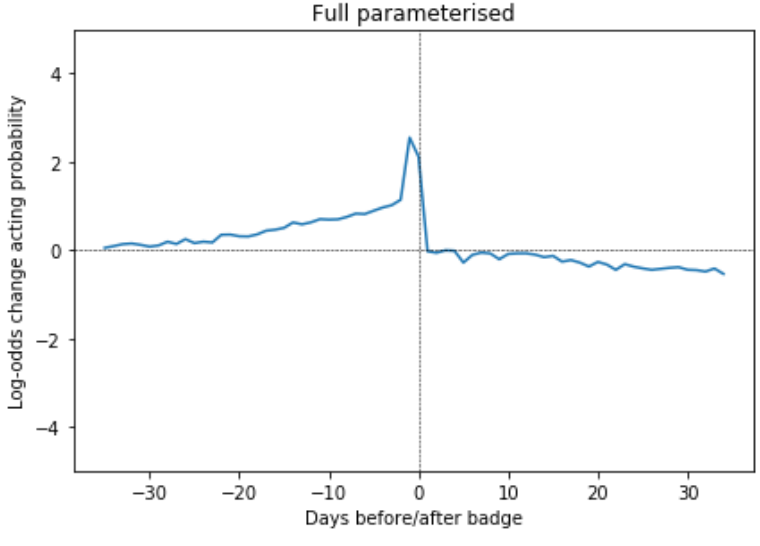
Average Test loss: 38.5821

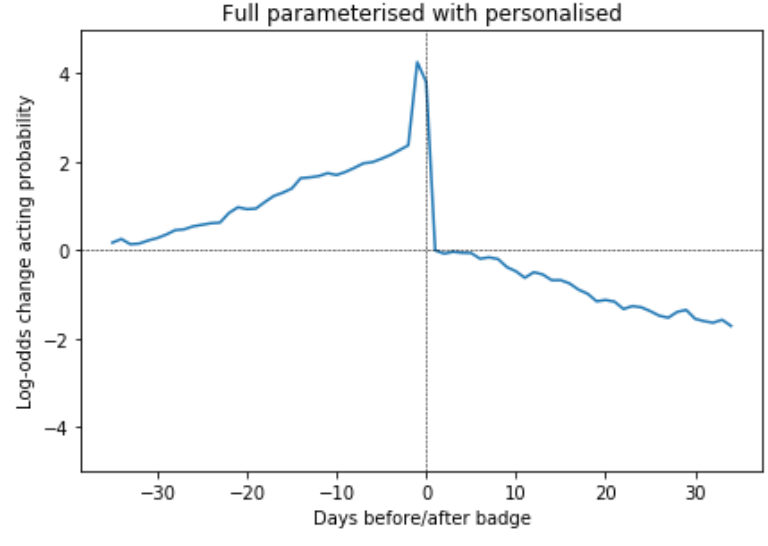

Histogram of the latent space that controls the steering param
entire population
not steered
steered
50
40
30
20
10
0
−4
−3
−2
−1
0
1
2
3

Compare the results:


Linear
Log-odds change acting probability
Days before/after badge
4
2
0
−2
−4
−30
−20
−10
0
10
20
30


Linear with personalised
Log-odds change acting probability
Days before/after badge
4
2
0
−2
−4
−30
−20
−10
0
10
20
30

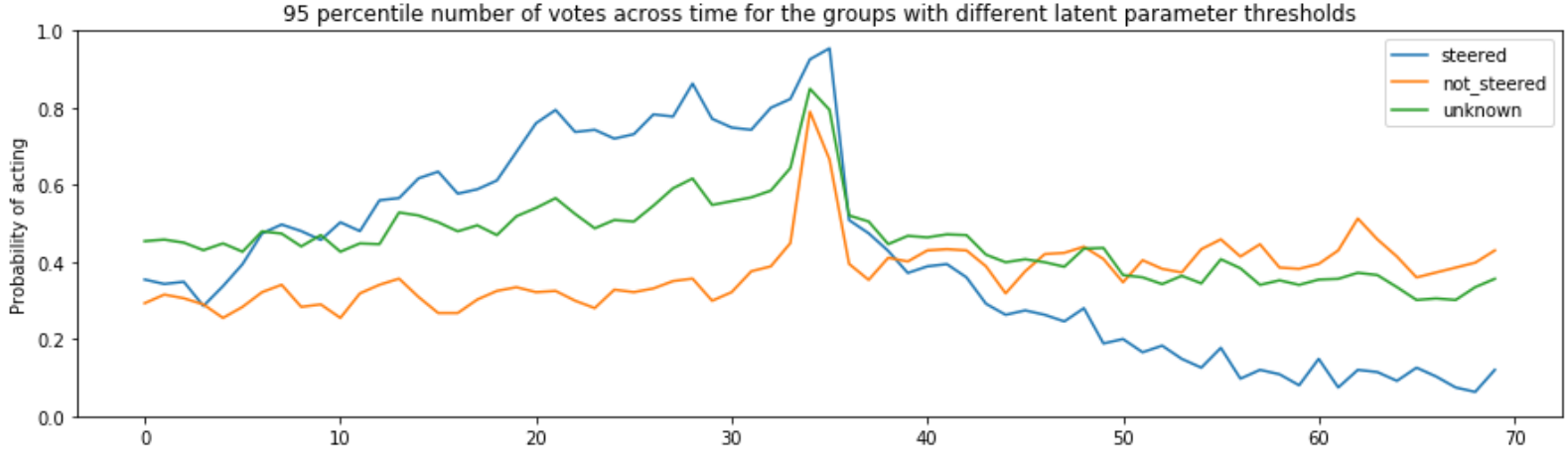

Full parameterised
Log-odds change acting probability
Days before/after badge
4
2
0
−2
−4
−30
−20
−10
0
10
20
30


Full parameterised with personalised
Log-odds change acting probability
Days before/after badge
4
2
0
−2
−4
−30
−20
−10
0
10
20
30

```
Baseline:                Average Test loss: 40.8441
Linear:                  Average Test loss: 40.0458
Linear + Personalised:   Average Test loss: 39.2855
Full P:                  Average Test loss: 39.0970
Full P + Personalised:   Average Test loss: 38.5056
```

Next Steps

Study the trajectories of the inferences from Full-parameterised model:


95 percentile number of votes across time for the groups with different latent parameter thresholds
Probability of acting
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
steered
not_steered
unknown