

Maximum Likelihood Estimation for SOF

Andrea Stocco, University of Washington
stocco@uw.edu

Goals

- Using Maximum Likelihood Estimation to find the most likely values of SOF (and other parameters) from the MemoryLab lesson data
- Make the SOF results more reliable and reproducible

What is MLE?

Find the parameters θ of a model that maximize its likelihood \mathcal{L}

$$\mathcal{L}(m, \theta | x) = P(x | m, \theta)$$

model ↓
parameters data

In practice, you use **log-likelihood**, because probs become vanishingly small when there are series of products

$$\log \mathcal{L}(m, \theta | x) = \log P(x | m, \theta)$$

Trial by trial likelihood

$$\mathcal{L}(m, \theta | x) = P(x | m, \theta); \quad x = \{x_1, x_2, \dots x_N\}$$

$$P(x|m, \theta) = P(x_1|m, \theta) \cdot P(x_2|m, \theta, x_1) \cdot \dots \cdot P(x_N|m, \theta, x_1, x_2 \dots x_{N-1})$$

ACT-R is a Markov model, and every choice is determined only by the current state.

So, if we force the model to follow the choices:

$$P(x|m, \theta) = P(x_1|m, \theta) \cdot P(x_2|m, \theta) \cdot \dots \cdot P(x_N|m, \theta)$$

$$\log \mathcal{L} = \sum_i \log P(x_i|m, \theta)$$

This is just **model tracing!** (Koedinger & Anderson, 1993)

The main equations

- The law of memory: Sum of fading traces + noise (s)

$$A(m, t) = \sum_i (t - t_i)^{-d(i)} + s$$

- Fading of traces depends on memory activation + fixed part (a)

$$d(i) = c e^{-A(m, t)} + a$$

- Response times depend on activation

$$RT = T_{ER} + F e^{-A(m, t)}$$

Probability density functions

- To apply MLE, we need to know the **probability density functions** (PDFs) of observed behaviors, given the model parameters
- We need two PDFs, one for **accuracies** and one for **RTs**

Probability density functions: Accuracies

- Accuracies are easy! The probability of a correct response is just the probability of retrieving memory m :

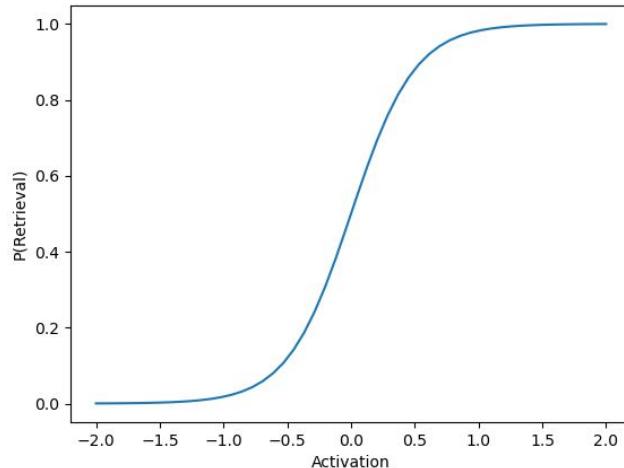
$$P(m) = 1 / (1 + e^{[-A(m) - \tau]/s})$$

- Where:
 - τ = retrieval threshold
 - s = activation noise

Probability density functions: Accuracies

The probability of retrieving memory m as a function activation $A(m)$

$$P(m) = 1 / (1 + e^{[-A(m) - \tau]/s})$$



Probability density function: Response Times

The probability of a particular response time RT when retrieving m depends on the distribution of activation noise s .

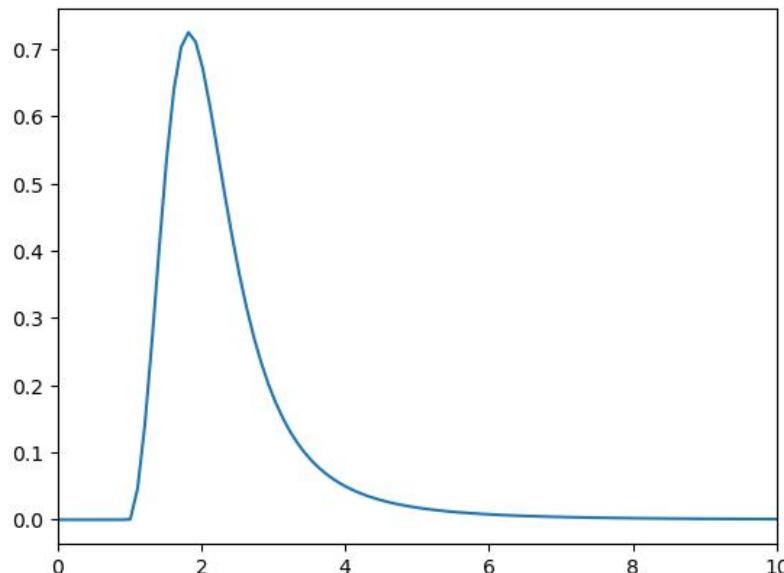
Noise is distributed according to a **logistic distribution**; therefore, the activation $A(m) + s$ is distributed as a logistic distribution with mean $\mu = A(m)$ and $\sigma = \pi s / \sqrt{3}$.

That means, $T_{ER} + Fe^{A(m) + s}$ is a **shifted log-logistic** distribution with parameters $\alpha = e^\mu = e^{A(m)}$, $\beta = 1/\sigma = \sqrt{3}/\pi s$, and $\gamma = T_{ER}$

Probability density function /2

$$\alpha = e^{-A(m)}, \quad \beta = \sqrt{3/\pi s}, \quad \gamma = T_{ER}; \quad T_{ER} = 1.0 \text{ s}, \quad F = 1.0$$

$$P(t) = (\beta/F\alpha)((t - T_{ER})/\alpha)^{\beta-1} / (1 + (t - T_{ER})/(F\alpha)^\beta)^2$$



Next Steps

Step 1: Simple MLE on accuracies and RTs

- Use standard algorithms (e.g., Powell's method) to maximize likelihood for a series of accuracies and response times.
- The main parameters are “AFTER”: α, F, T_{ER}
 - All AFTER params are at the subject-level
- In addition, a constant $C_{f_1}, C_{f_2}, \dots, C_{f_N}$ is estimated for each fact f_1, f_2, \dots, f_N .
 - The constants are fact-level, and each session will have a different N
- The constant is added to each fact's activation (:BLC) during the probability density estimation

Step 2

Use **Expectation-Maximization**. Apply MLE recursively between AFTER parameters and fact-specific constants C_1, C_2, \dots, C_N

$$C_1, C_2, \dots, C_N = 0$$



Fix constants, estimate
AFTER



Fix AFTER,
Estimate constants



Until C_1, C_2, \dots, C_N are **stable**

Step 3: Bayesian Priors

Start with priors for the AFTER parameters

Instead of MLE, use MAP (Maximum A Priori):

$$\theta^* = \operatorname{argmax}_{\theta} \log P(x | m, \theta) P(\theta) = \operatorname{argmax}_{\theta} \mathcal{L} P(\theta)$$