Building a Model: Maximum Likelihood Estimation (MLE)

Katharina V. Wellstein

nature
computational psychiatry

ARTICLES

How to get rich: our fancy model for maximizing wins at the slot machine

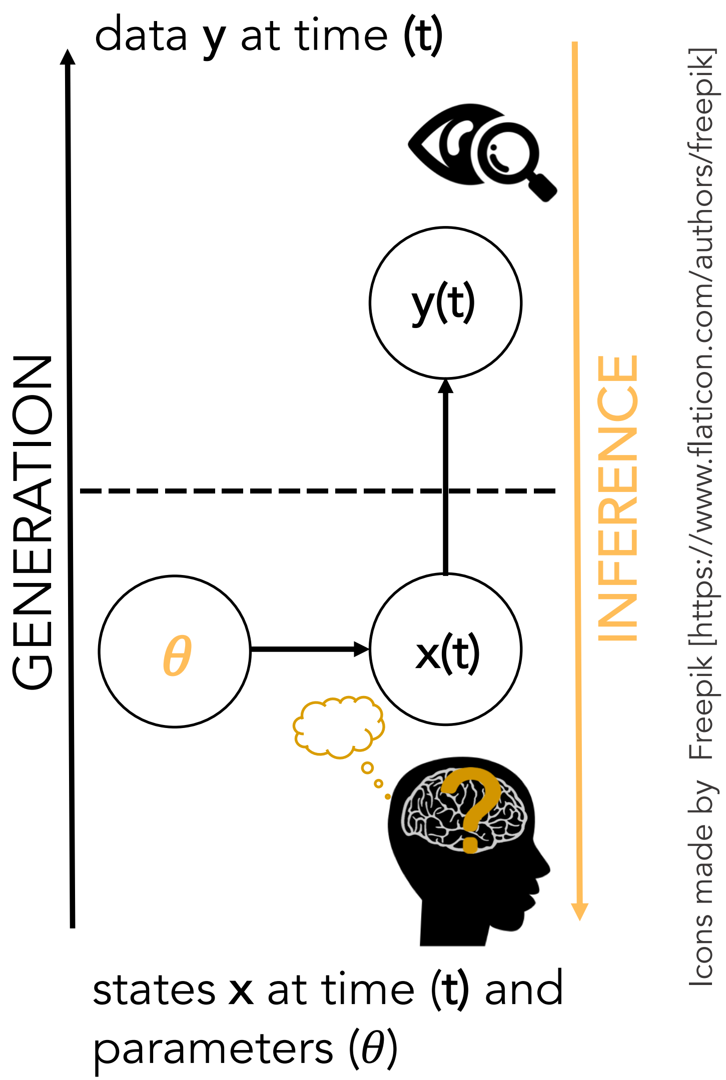
Your Name [1,2]

Casinos are an attractive alternative to third-party funding for starting your own research group or running projects you have in mind. We here present a fancy computational model that allows you to maximize the money you can get from playing the slot machine at your local casino. This will allow you to build your own group and run all the awesome experiments that you were planning for so long but never had the financial resources to do. You might even be able to invest in this awesome wide-screen monitor and the Italian coffee machine that you always wanted to have in your lab. Hence, this paper is of high practical relevance for the scientific community

REMINDER: STEPS TAKEN SO FAR

- ✓ Specified plausible models
- ✓ Simulated data
- ✓ Acquired data

? Inference / Model Inversion: Acquired Data ← Parameters



REMINDER: OUR MODEL SPACE

or what models may describe successful slot-machine players' behavior

model 1
Random choice

$$p_t^1 = b \qquad p_t^0 = 1 - b$$

$$0 \le b \le 1$$

$$\boldsymbol{\theta} = \{\boldsymbol{b}\}$$

model 2
Noisy win-stay-lose-switch

$$p_t^k = \begin{cases} 1 - \frac{\varepsilon}{2} & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \frac{\varepsilon}{2} & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$

$$\boldsymbol{\theta} = \{\boldsymbol{\varepsilon}\}$$

model 3
Rescorla Wagner

$$Q_{t+1}^k = Q_t^k + \alpha\left(r_t - Q_t^k\right) \quad \text{and} \quad p_t^k = \frac{\exp\left(\beta Q_t^k\right)}{\sum_{i=1}^K \exp\left(\beta Q_t^i\right)}$$

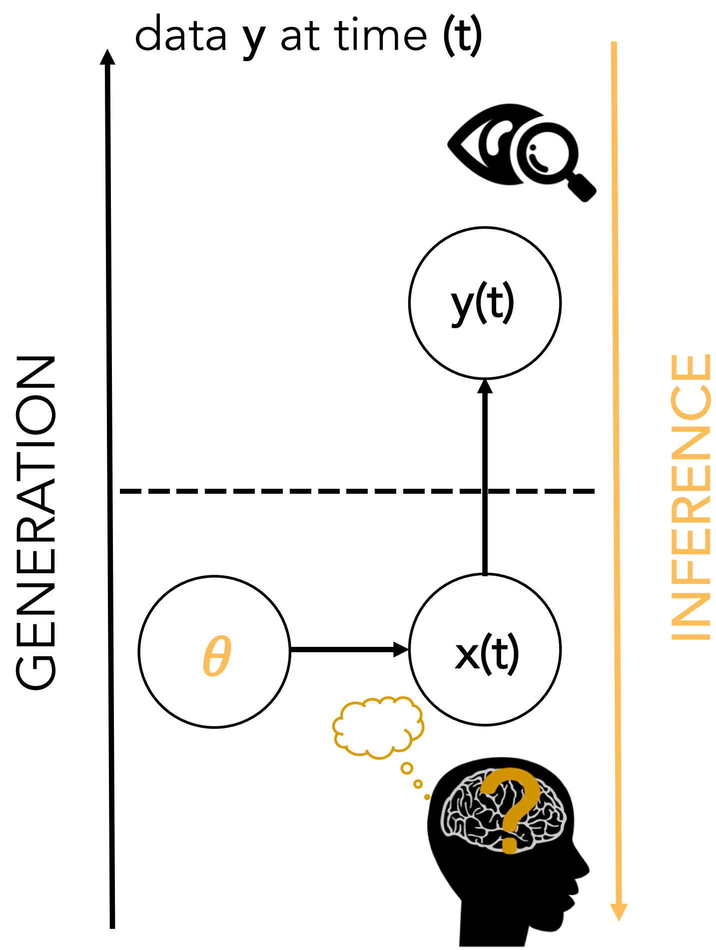
$$\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$$

BASIS OF LME : ESTIMATOR FUNCTION

$$\theta_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}}\big(p(Y|\theta, m)\big)$$, where:

$$p(Y|\theta, m) = p(y_{1..T}|\theta, m)$$ Likelihood

"Maximum likelihood estimation finds the θ for which the acquired data is most likely under the model"



Icons made by Freepik [https://www.flaticon.com/authors/freepik]

SPECIFYING THE LIKELIHOOD FUNCTION

model 1
Random choice

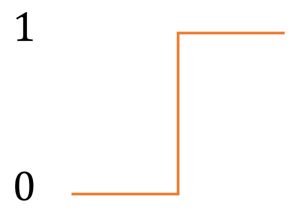
$p_t^1 = b$

$p_t^0 = 1 - b$

$0 \leq b \leq 1$

$\boldsymbol{\theta} = \{\boldsymbol{b}\}$

For single trial t: $p(y_t|\theta, m) = \theta^{y_t}(1-\theta)^{(1-y_t)}$ Bernoulli distribution



$b^0 = 1$

For all trials $1..T$: $p(Y|\theta, m) = p(y_{1..T}|\theta, m) = \prod_{t=1}^{T} \theta^{y_t}(1-\theta)^{(1-y_t)}$

$Y = \{y_1, \dots, y_T\}$ iid

MAXIMISING THE LIKELIHOOD FUNCTION

$$p(Y|\theta, m) = \prod_{t=1}^{T} p(y_t|\theta, m)$$

Likelihood

Analytical Solution

Is $p(Y|\theta, m)$ differentiable?
Is it tractable?

→ use the derivative test,
i.e. set the derivative to 0.

In simple cases possible to
find the maximum analytically

Numerical Solution

ANALYTICAL SOLUTION: THE LIKELIHOOD FUNCTION

model 1
Random choice

$$p(Y|\theta, m) = \prod_{t=1}^{T} \theta^{y_t}(1-\theta)^{(1-y_t)}$$

Logarithm & Exponential

$y = \log_b(X) \;\; -> b^y = X$

$$\log\big(p(Y|\theta, m)\big) = \log\left(\prod_{t=1}^{T} \theta^{y_t}(1-\theta)^{(1-y_t)}\right)$$

Product Rule for Logarithms

$\log(X_1 X_2) = \log(X_1) + \log(X_2) \; for \; b > 0$

$$= \sum_{t=1}^{T} \log\big(\theta^{y_t}(1-\theta)^{(1-y_t)}\big)$$

Logarithm of a power

$log\left((X_1)^{(X_2)}\right) = X_2 \log(X_1) \; for \; b > 0$

$$= \sum_{t=1}^{T} \big(y_t \log(\theta) + (1-y_t)\log(1-\theta)\big)$$

$$\frac{d}{d\theta}\log\big(p(Y|\theta, m)\big) = \frac{d}{d\theta}\sum_{t=1}^{T} \big(y_t \log(\theta) + (1-y_t)\log(1-\theta)\big) \overset{!}{=} 0$$

ANALYTICAL SOLUTION: THE LIKELIHOOD FUNCTION

model 1
Random
choice

$$\frac{d}{d\theta}\sum_{t=1}^{T}(y_t\log(\theta)+(1-y_t)\log(1-\theta))=0$$

$$\left(\frac{d}{d\theta}\log(\theta)\right)\left(\sum_{t=1}^{T}y_t\right)+\left(\frac{d}{d\theta}\log(1-\theta)\right)\left(\sum_{t=1}^{T}(1-y_t)\right)=0$$

$$\frac{1}{\theta(1-\theta)}\left(\sum_{t=1}^{T}y_t-\theta\sum_{t=1}^{T}y_t-\theta T+\theta\sum_{t=1}^{T}y_t\right)=0$$

$$\sum_{t=1}^{T}y_t-\theta T=0$$

$$\theta_{MLE}=\frac{1}{T}\sum_{t=1}^{T}y_t$$

MAXIMISING THE LIKELIHOOD FUNCTION

$$p(Y|\theta, m) = \prod_{t=1}^{T} p(y_t|\theta, m)$$

Likelihood

Analytical Solution

Is $p(Y|\theta, m)$ is differentiable?

→ use the derivative test, i.e. set the derivative to 0.

In simple cases possible to find the maximum analytically

Numerical Solution

→ use numerical routines available in different software.

e.g. fminsearch* or fmincon** in MATLAB

*"Find minimum of unconstrained multivariable function using derivative-free method"

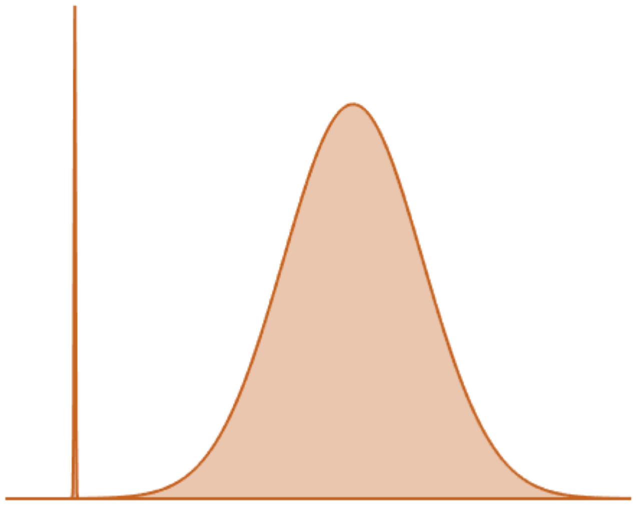
**" Find minimum of constrained nonlinear multivariable function"

ADVANTAGES OF MLE

- Easy to compute
- Interpretable
- Asymptotic properties
- Invariant to reparameterization: if θ_{MLE} is a MLE for θ, then $g(\theta_{\mathrm{MLE}})$ is a MLE for $g(\theta)$

LIMITATIONS OF MLE

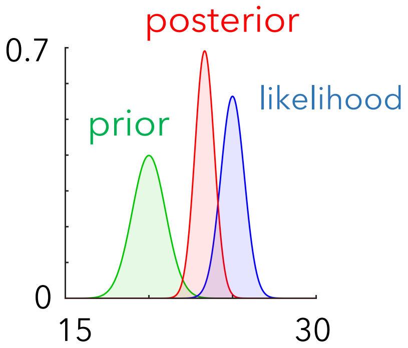
- Existence & uniqueness of ML Estimator is not guaranteed
- ML Estimator is a point estimate and therefore has no representation of uncertainty
- ML Estimator might not be representative
- Overfitting



ALTERNATIVES FOR MLE

Bayesian statistics

$$\underbrace{p(\theta|y,m)}_{\text{posterior}} = \frac{\overbrace{p(y|\theta,m)}^{\text{likelihood}}\overbrace{p(\theta|m)}^{\text{prior}}}{\underbrace{p(y|m)}_{\text{model evidence}}}$$



- maximum-a-posteriori (MAP) estimation:
  → Point estimates of the posterior
  ⚠️ Under a flat prior, MAP = MLE

- Variational Bayesian (VB) or sampling-based (Markov Chain Monte Carlo) techniques
  → Full posterior densities

THANK YOU FOR YOUR ATTENTION

Katharina V. Wellstein

✉️ wellstein@biomed.ee.ethz.ch

@kv_wellstein

*Twitter Icon: https://www.flaticon.com/free-icons/twitter, created by Pixel perfect - Flaticon