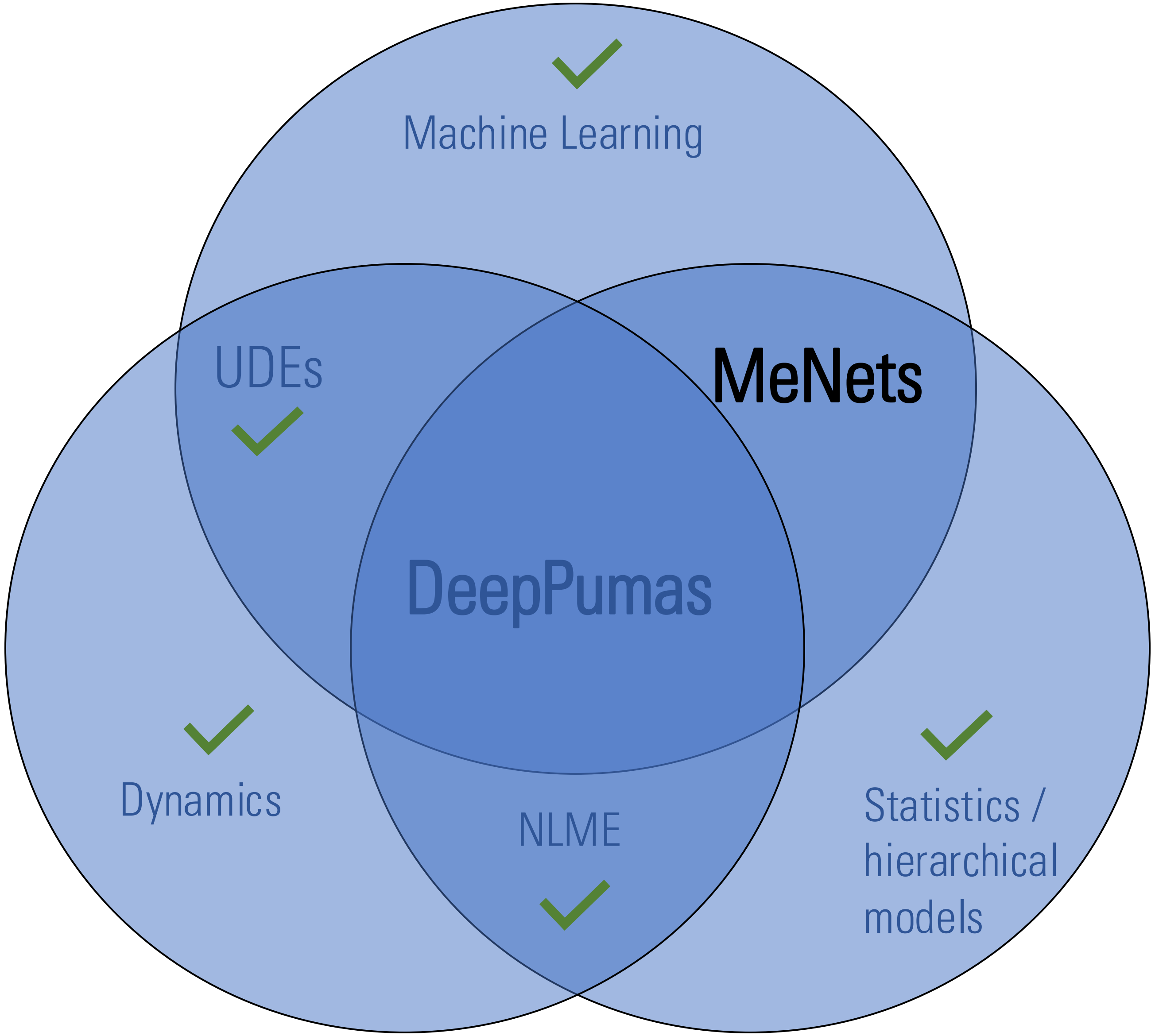
DeepPumas Embedded ML

Niklas Korsbo and Mohamed Tarek

DeepPumas



Let's have a look at MeNets.

The integration of mixed effects and neural networks.

What are Mixed Effects?

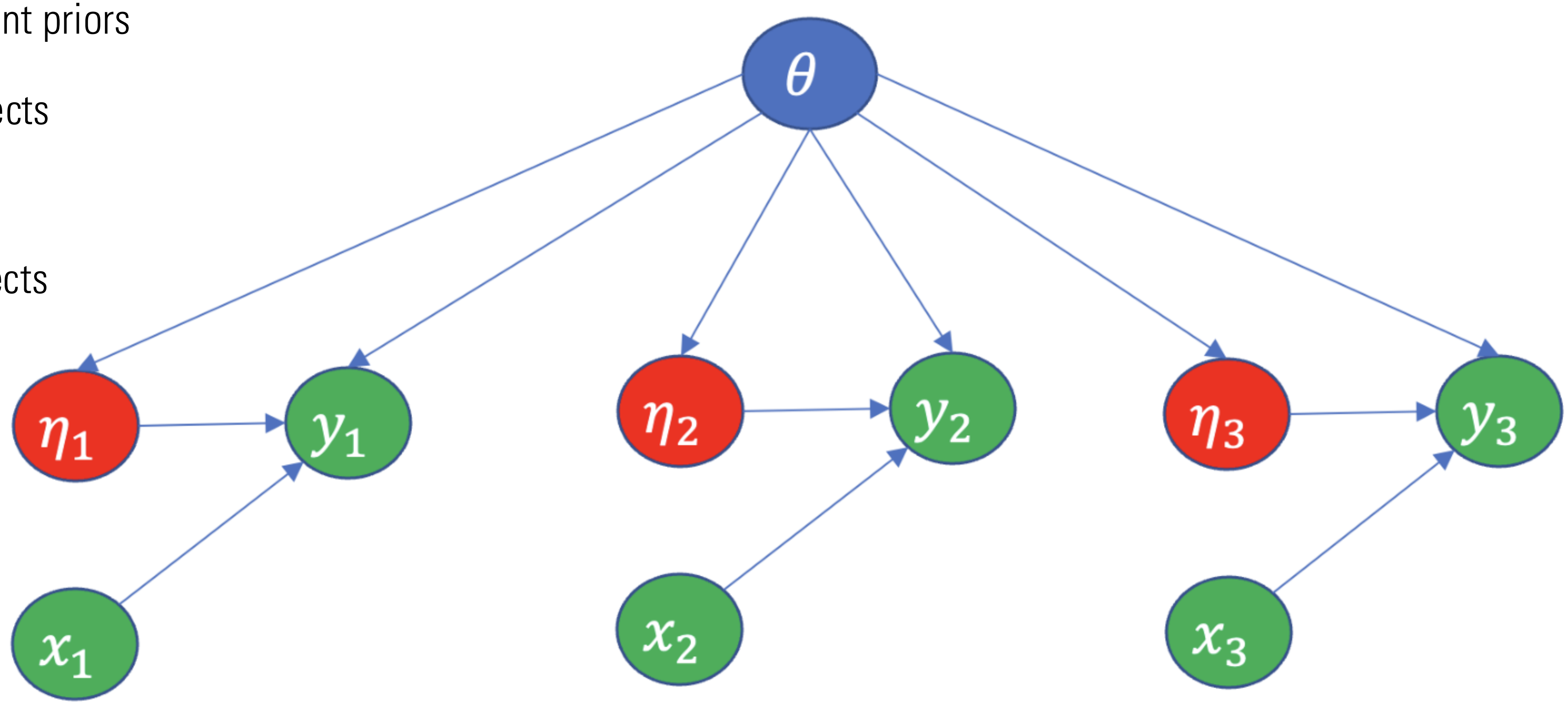
- Fixed effects, θ
    - Model parameters modelled as deterministic quantities
- Random effects, η
    - Model parameters modelled as random variables

Hierarchical

We typically define hierarchies where θ are shared parameters but η is subject-specific.

DeepPumas

- θ – all the population parameters.
  - Shared across subjects
- η – random effects of all subjects.
  - η_1 specific to subject 1
  - Typically has θ-dependent priors
- x – covariates of all the subjects
  - x_1 specific to subject 1
- y – responses of all the subjects
  - y_1 specific to subject 1



DeepPumas

In a Pumas model

θ

$\eta_i \mid \theta$

x_i

$y_i | \theta, \eta_i, x_i$

```
@model begin
  @param begin
    θ ∈ VectorDomain(4, lower = zeros(4))
    Ω ∈ PSDDomain(2)
    Σ ∈ RealDomain(lower = 0.0)
    a ∈ RealDomain(lower = 0.0, upper = 1.0)
  end
  @random begin
    η ~ MvNormal(Ω)
  end
  @covariates sex wt etn
  @pre begin
    θ1 := θ[1]
    Ka = θ1
    CL = θ[2] * ((wt / 70)^0.75) * (θ[4]^sex) *
      exp(η[1])
    Vc = θ[3] * exp(η[2])
  end
  @dynamics begin
    Depot' = -Ka * Depot
    Central' = Ka * Depot - (CL / Vc) * Central
    Res' = Depot - Central
  end
  @derived begin
    conc = @. Central / Vc
    dv ~ @. Normal(conc, conc * Σ)
    T_max = maximum(t)
  end
  @observed begin
    obs_cmax = maximum(dv)
  end
end
```

DeepPumas

Don’t assign too much meaning to the random effects

- Indicates unknown parameters that vary between subjects (or whatever hierarchy we use)
- Usually tied very closely to a fixed effect in pharmacometrics. $CL = tvCL \cdot \exp(\eta_{cl})$.
- Used in machine learning (called “latent variables”) without assigning much meaning to them.
- Enables degrees of freedom along which the model can account for outcome heterogeneity.

DeepPumas

Random effects during simulation?

Simple

Just sample and use

θ

$\eta_i \mid \theta$

x_i

$y_i|\theta, \eta_i, x_i$

```
@model begin
  @param begin
    θ ∈ VectorDomain(4, lower = zeros(4))
    Ω ∈ PSDDomain(2)
    Σ ∈ RealDomain(lower = 0.0)
    a ∈ RealDomain(lower = 0.0, upper = 1.0)
  end
  @random begin
    η ~ MvNormal(Ω)
  end
  @covariates sex wt etn
  @pre begin
    θ1 := θ[1]
    Ka = θ1
    CL = θ[2] * ((wt / 70)^0.75) * (θ[4]^sex) *
      exp(η[1])
    Vc = θ[3] * exp(η[2])
  end
  @dynamics begin
    Depot' = -Ka * Depot
    Central' = Ka * Depot - (CL / Vc) * Central
    Res' = Depot - Central
  end
  @derived begin
    conc = @. Central / Vc
    dv ~ @. Normal(conc, conc * Σ)
    T_max = maximum(t)
  end
  @observed begin
    obs_cmax = maximum(dv)
  end
end
```

DeepPumas

Fitting with random effects

Their effect is largely determined by how they contribute to the loss function of a model fit

Conditional probability

Probability of the response $\boldsymbol{y}$ according to the model given specific values of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{x}$

$$p_c(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{x})$$

Fit model by simply finding the values of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ that maximizes the conditional probability?

DeepPumas

Fitting with random effects

Their effect is largely determined by how they contribute to the loss function of a model fit

<u>Marginal probability (!)</u>

Integrates out the effect of the random effects

$$p_m(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{x}) = \int p_c(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{x}) \cdot p_{prior}(\boldsymbol{\eta} \mid \boldsymbol{\theta})\, d\boldsymbol{\eta}$$

Average conditional probability weighted by a prior.

Different methods/approximations: Laplace, FOCE and EM

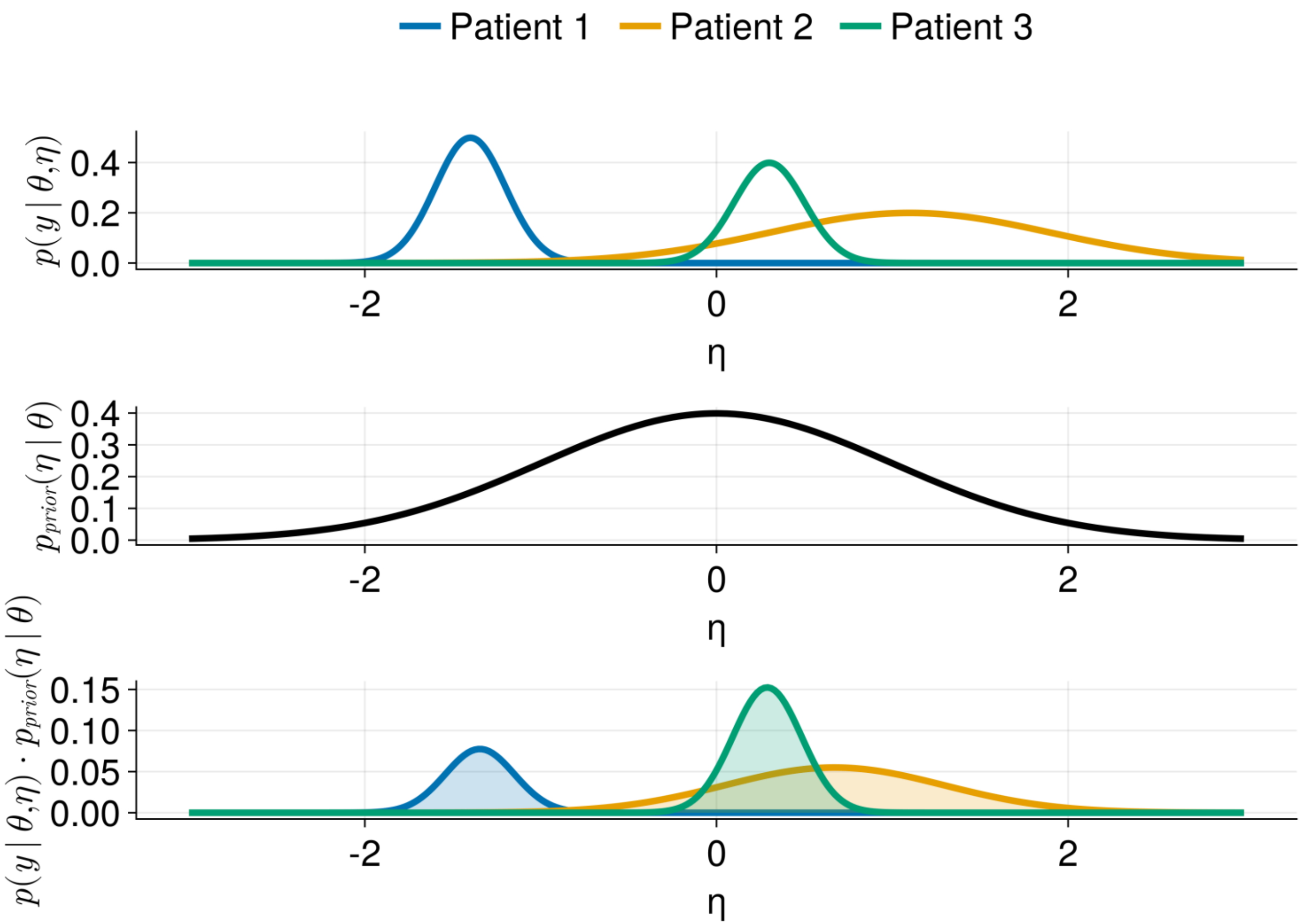
DeepPumas

Fitting with random effects

Their effect is largely determined by how they contribute to the loss function of a model fit

Conditional likelihood

Marginal likelihood



DeepPumas

Marginal vs conditional

- Consider a single subject i. The marginal likelihood is:

$$p(\boldsymbol{y_i} \mid \boldsymbol{\theta}) = \int p(\boldsymbol{y_i} \mid \boldsymbol{\eta_i}, \boldsymbol{\theta}) \cdot p(\boldsymbol{\eta_i} \mid \boldsymbol{\theta})\, d\boldsymbol{\eta_i}$$

- However, we can write the marginal likelihood in another way:

$$p(\boldsymbol{y_i} \mid \boldsymbol{\theta}) = \prod_{j=1}^{m_i} p(y_{i,j} \mid \boldsymbol{y_{i,1:j-1}}, \boldsymbol{\theta})$$

where $\boldsymbol{y_{i,1:j}}$ are the observations of subject i until time point t_j, and j is an integer that goes from 1 to m_i (number of longitudinal observations for subject i).

DeepPumas

Marginal vs conditional

- No past observations, $j = 1$

$$p\left(y_{i,1} \mid \boldsymbol{y}_{\boldsymbol{i},\mathbf{1:0}}, \boldsymbol{\theta} \right) = p\left(y_{i,1} \mid \boldsymbol{\theta} \right) = \int p\left(y_{i,1} \mid \boldsymbol{\eta}_{\boldsymbol{i}}, \boldsymbol{\theta} \right) \cdot p\left(\boldsymbol{\eta}_{\boldsymbol{i}} \mid \boldsymbol{\theta} \right) d\boldsymbol{\eta}_{\boldsymbol{i}}$$

- With past observations, $j > 1$

$$p\left(y_{i,j} \mid \boldsymbol{y}_{\boldsymbol{i},\mathbf{1}:\boldsymbol{j}-\mathbf{1}}, \boldsymbol{\theta} \right) = \int p\left(y_{i,j} \mid \boldsymbol{\eta}_{\boldsymbol{i}}, \boldsymbol{\theta} \right) \cdot p\left(\eta_i \mid \boldsymbol{y}_{\boldsymbol{i},\mathbf{1}:\boldsymbol{j}-\mathbf{1}}, \boldsymbol{\theta} \right) d\boldsymbol{\eta}_{\boldsymbol{i}}$$

where $p(\boldsymbol{\eta}_{\boldsymbol{i}} \mid \boldsymbol{y}_{\boldsymbol{i},\mathbf{1}:\boldsymbol{j}-\mathbf{1}}, \boldsymbol{\theta})$ is the posterior distribution of $\boldsymbol{\eta}_{\boldsymbol{i}}$ given partial data $\boldsymbol{y}_{\boldsymbol{i},\mathbf{1}:\boldsymbol{j}-\mathbf{1}}$ and $\boldsymbol{\theta}$.

DeepPumas

Fitting algorithms in DeepPumas

	Marginal likelihood	Conditional likelihood
Prior / regularization on θ	MAP(FO()) MAP(FOCE()) MAP(LaplaceI())	JointMAP()
No prior / regularization on θ	FO() FOCE() LaplaceI() SAEM()	N/A

DeepPumas

Algorithms

- FO(), FOCE(), LaplaceI() and SAEM()

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\arg\max} \prod_{i=1}^{N} p(\boldsymbol{y_i} \mid \boldsymbol{\theta}, \boldsymbol{x})$$

$$\boldsymbol{\eta}_{\boldsymbol{i}}^* = \underset{\boldsymbol{\eta}_i}{\arg\max}\ p(\boldsymbol{y_i} \mid \boldsymbol{\theta} = \boldsymbol{\theta}^*, \boldsymbol{\eta}_i, \boldsymbol{x_i}) \cdot p(\boldsymbol{\eta_i} \mid \boldsymbol{\theta} = \boldsymbol{\theta}^*)$$

- MAP(FO()), MAP(FOCE()) and MAP(LaplaceI())

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\arg\max}\ p(\boldsymbol{\theta}) \cdot \prod_{i=1}^{N} p(\boldsymbol{y_i} \mid \boldsymbol{\theta}, \boldsymbol{x})$$

$$\boldsymbol{\eta}_{\boldsymbol{i}}^* = \underset{\boldsymbol{\eta}_i}{\arg\max}\ p(\boldsymbol{y_i} \mid \boldsymbol{\theta} = \boldsymbol{\theta}^*, \boldsymbol{\eta}_i, \boldsymbol{x_i}) \cdot p(\boldsymbol{\eta_i} \mid \boldsymbol{\theta} = \boldsymbol{\theta}^*)$$

Algorithms

- JointMAP()

$$
\begin{aligned}
\boldsymbol{\theta}^*, \boldsymbol{\eta}^* &= \arg\max_{(\boldsymbol{\theta},\boldsymbol{\eta})} p(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{x}, \boldsymbol{y}) \\
&= \arg\max_{(\boldsymbol{\theta},\boldsymbol{\eta})} p(\boldsymbol{\theta}) \cdot \prod_{i=1}^{N} p(\boldsymbol{\eta}_{\boldsymbol{i}} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{y}_{\boldsymbol{i}} \mid \boldsymbol{\theta}, \boldsymbol{\eta}_{\boldsymbol{i}}, \boldsymbol{x}_{\boldsymbol{i}})
\end{aligned}
$$

- BayesMCMC()
  - Samples from the joint posterior $p(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{x}, \boldsymbol{y})$

Mixed effect neural networks (MeNets)

- What happens when you use random effects as part of your NN input?
- Let's see in an exercise!