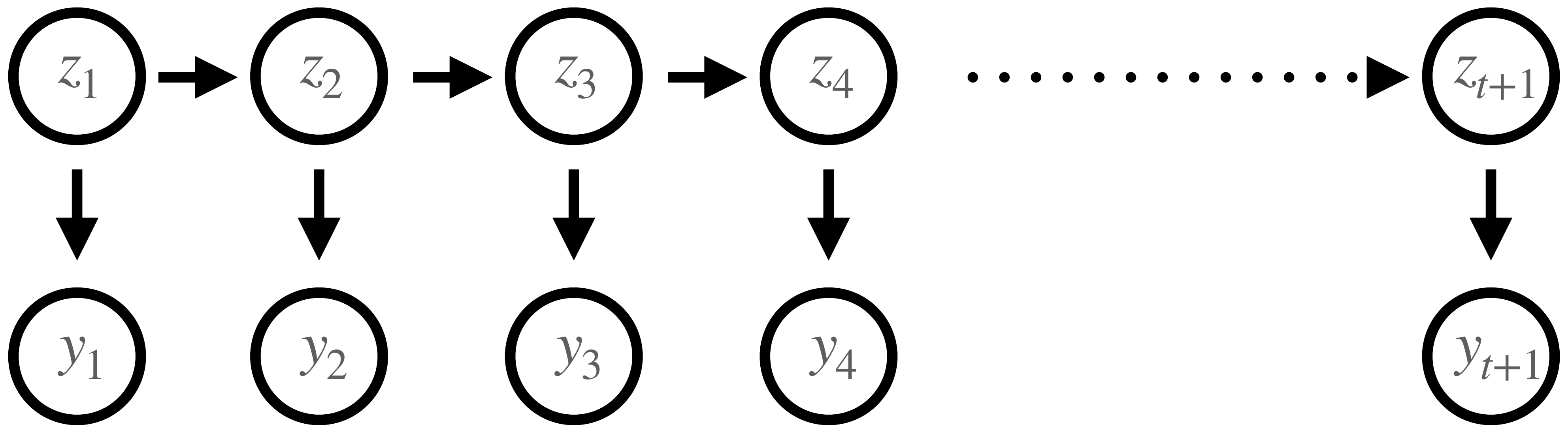
Extensions to LDS/HMM Models

+

Beginnings of Inference Techniques

Extensions to LDS/HMM Models

Linear Dynamical Systems



$$P(z_{1:T}, y_{1:T}) = P(z_1)\prod_{t=2}^{T} p(z_t \mid z_{t-1})\prod_{t=1}^{T} p(y_t \mid z_t)$$

$$\mathcal{N}(\boldsymbol{z}|\mathbf{0}, \mathbf{I})$$

$$\mathcal{N}(\boldsymbol{z}_t|\mathbf{F}_t\boldsymbol{z}_{t-1}, \mathbf{Q}_t)$$

$$\mathcal{N}(\boldsymbol{y}_t|\mathbf{H}_t\boldsymbol{z}_t, \mathbf{R}_t)$$

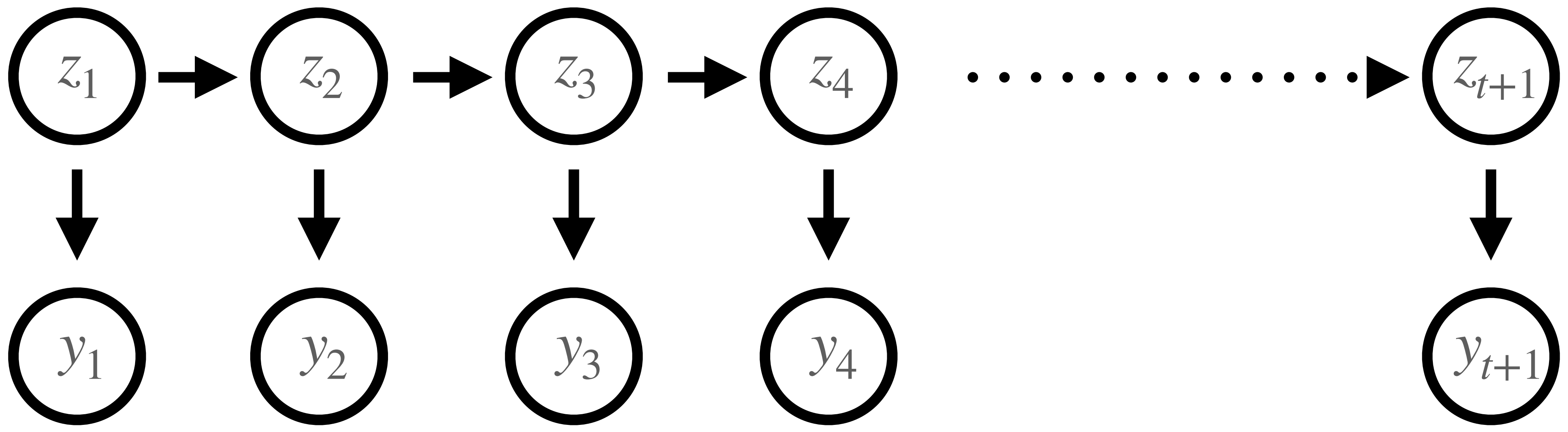
Or more fully:

$$\mathcal{N}(\boldsymbol{z}_t|\mathbf{F}_t\boldsymbol{z}_{t-1} + \mathbf{B}_t\boldsymbol{u}_t + \boldsymbol{b}_t, \mathbf{Q}_t)$$

Or more fully:

$$\mathcal{N}(\boldsymbol{y}_t|\mathbf{H}_t\boldsymbol{z}_t + \mathbf{D}_t\boldsymbol{u}_t + \boldsymbol{d}_t, \mathbf{R}_t)$$

Poisson Linear Dynamical Systems



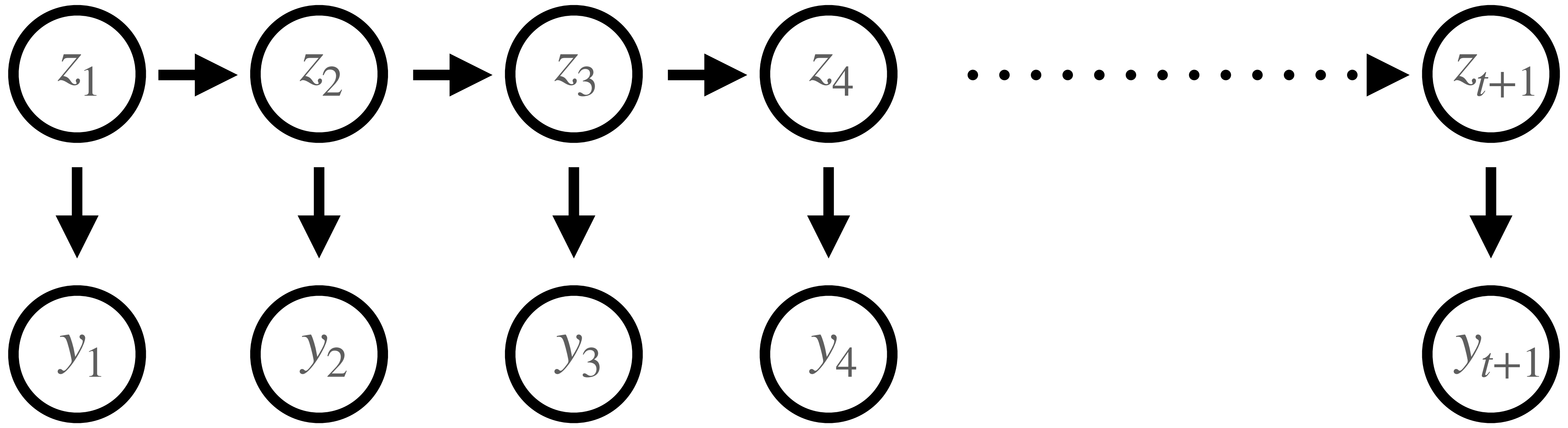
$$P(z_{1:T}, y_{1:T}) = P(z_1)\prod_{t=2}^{T} p(z_t \mid z_{t-1}) \prod_{t=1}^{T} p(y_t \mid z_t)$$

$$\mathcal{N}(\boldsymbol{z}|\mathbf{0}, \mathbf{I}) \qquad \mathcal{N}(\boldsymbol{z}_t|\mathbf{F}_t\boldsymbol{z}_{t-1}, \mathbf{Q}_t)$$

Or more fully:

$$\mathcal{N}(\boldsymbol{z}_t|\mathbf{F}_t\boldsymbol{z}_{t-1} + \mathbf{B}_t\boldsymbol{u}_t + \boldsymbol{b}_t, \mathbf{Q}_t)$$

Poisson Linear Dynamical Systems



$$P(z_{1:T}, y_{1:T}) = P(z_1) \prod_{t=2}^{T} p(z_t \mid z_{t-1}) \prod_{t=1}^{T} p(y_t \mid z_t)$$

f=exp() or softplus()

$$\mathcal{N}(\boldsymbol{z} | \mathbf{0}, \mathbf{I}) \qquad \mathcal{N}(\boldsymbol{z}_t | \mathbf{F}_t \boldsymbol{z}_{t-1}, \mathbf{Q}_t) \qquad Poisson\left(y_t \mid f(\mathbf{H}_t z_t)\right)$$

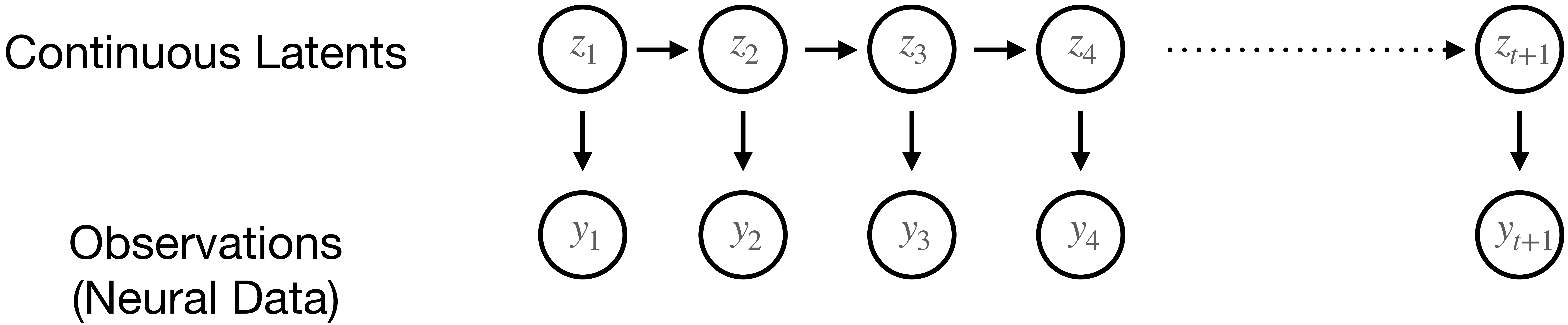
Or more fully:

$$\mathcal{N}(\boldsymbol{z}_t | \mathbf{F}_t \boldsymbol{z}_{t-1} + \mathbf{B}_t \boldsymbol{u}_t + \boldsymbol{b}_t, \mathbf{Q}_t)$$

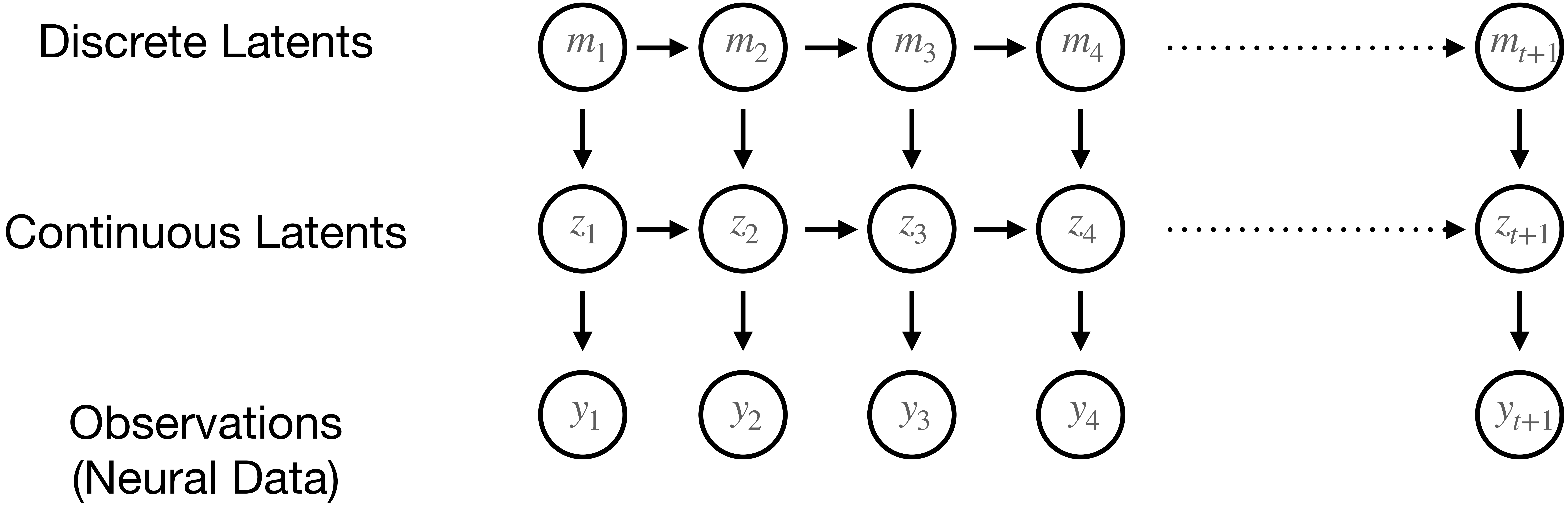
Or more fully:

$$Poisson\left(y_t \mid f(\mathbf{H}_t z_t + \mathbf{D}_t u_t + d_t)\right)$$

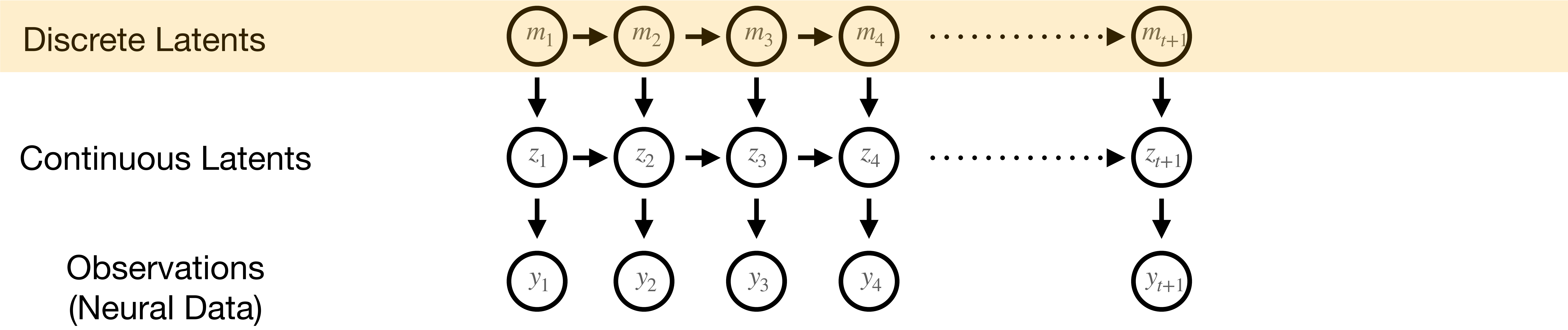
Switching Linear Dynamical Systems



Switching Linear Dynamical Systems



Switching Linear Dynamical Systems

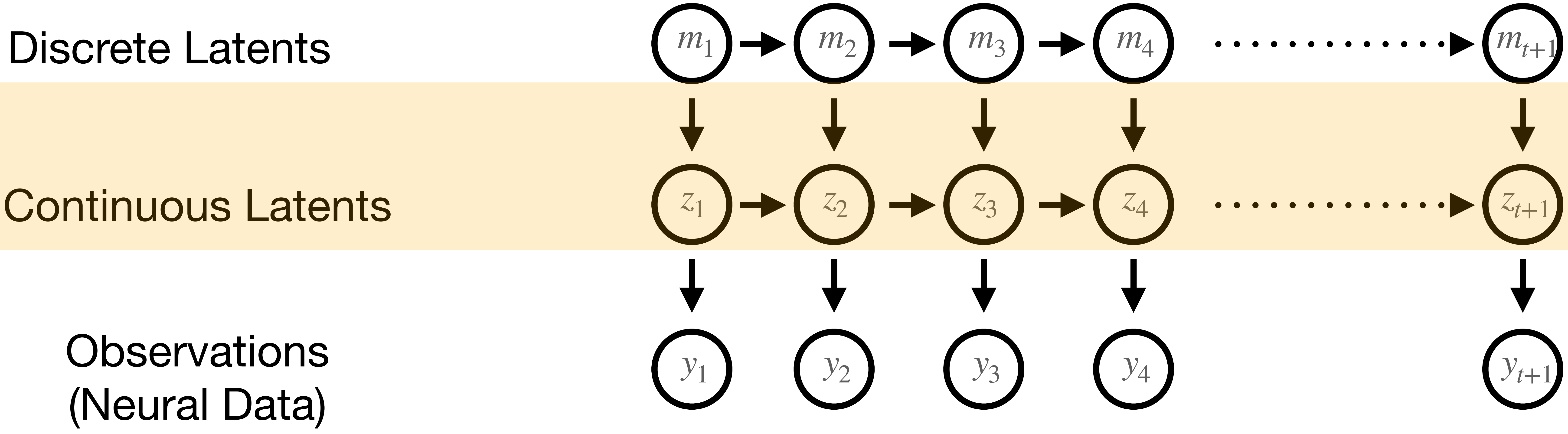


Discrete Latents $$p(m_t = k | m_{t-1} = j) = A_{jk}$$

Continuous Latents $$p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t | \mathbf{F}_k \boldsymbol{z}_{t-1} + \mathbf{B}_k \boldsymbol{u}_t + \boldsymbol{b}_k, \mathbf{Q}_k)$$

Observations (Neural Data) $$p(\boldsymbol{y}_t | \boldsymbol{z}_t, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t | \mathbf{H}_k \boldsymbol{z}_t + \mathbf{D}_k \boldsymbol{u}_t + \boldsymbol{d}_k, \mathbf{R}_k)$$

Switching Linear Dynamical Systems



Discrete Latents $p(m_t = k | m_{t-1} = j) = A_{jk}$

Continuous Latents $p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t | \mathbf{F}_k \boldsymbol{z}_{t-1} + \mathbf{B}_k \boldsymbol{u}_t + \boldsymbol{b}_k, \mathbf{Q}_k)$

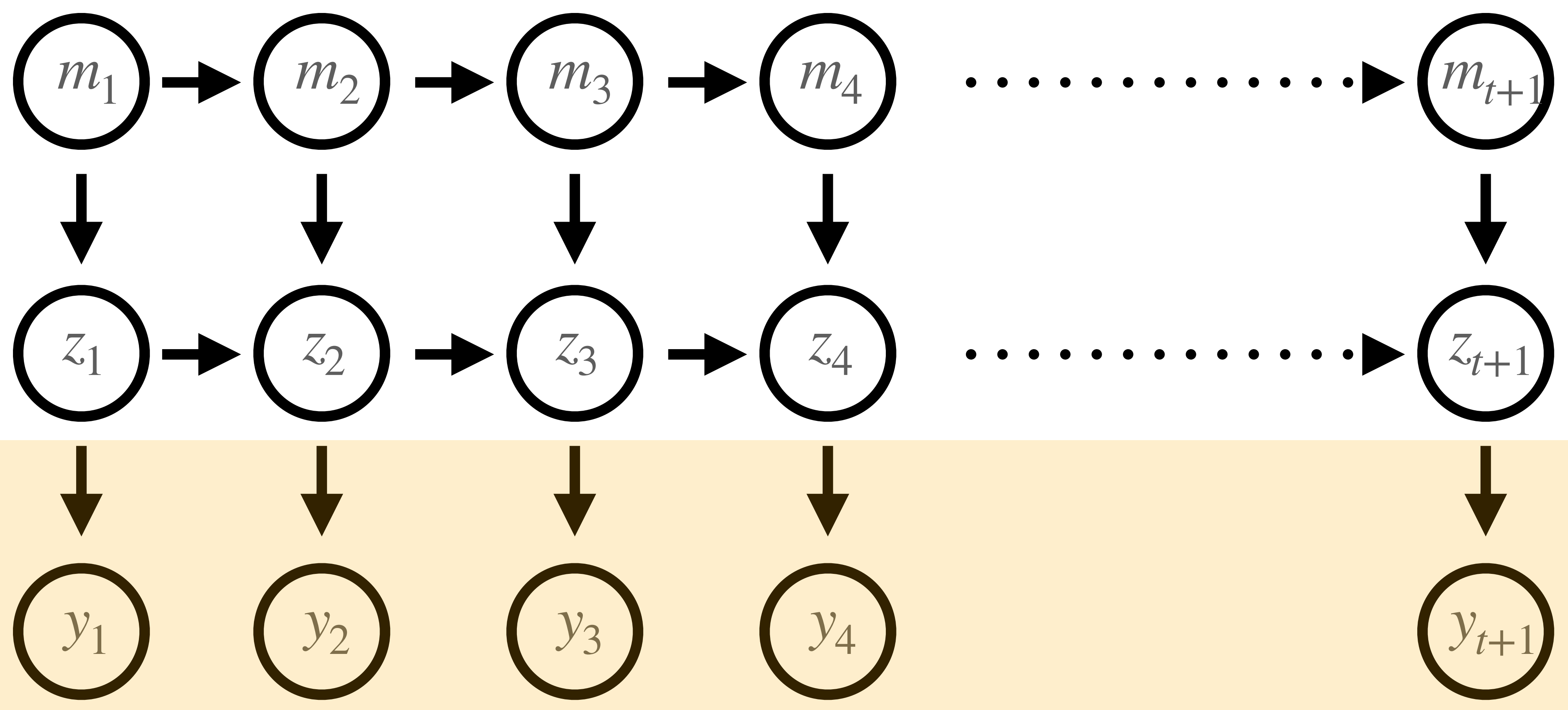
Observations (Neural Data) $p(\boldsymbol{y}_t | \boldsymbol{z}_t, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t | \mathbf{H}_k \boldsymbol{z}_t + \mathbf{D}_k \boldsymbol{u}_t + \boldsymbol{d}_k, \mathbf{R}_k)$

Switching Linear Dynamical Systems

Discrete Latents

Continuous Latents

Observations
(Neural Data)



Discrete Latents $p(m_t = k | m_{t-1} = j) = A_{jk}$

Continuous Latents $p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t | \mathbf{F}_k \boldsymbol{z}_{t-1} + \mathbf{B}_k \boldsymbol{u}_t + \boldsymbol{b}_k, \mathbf{Q}_k)$

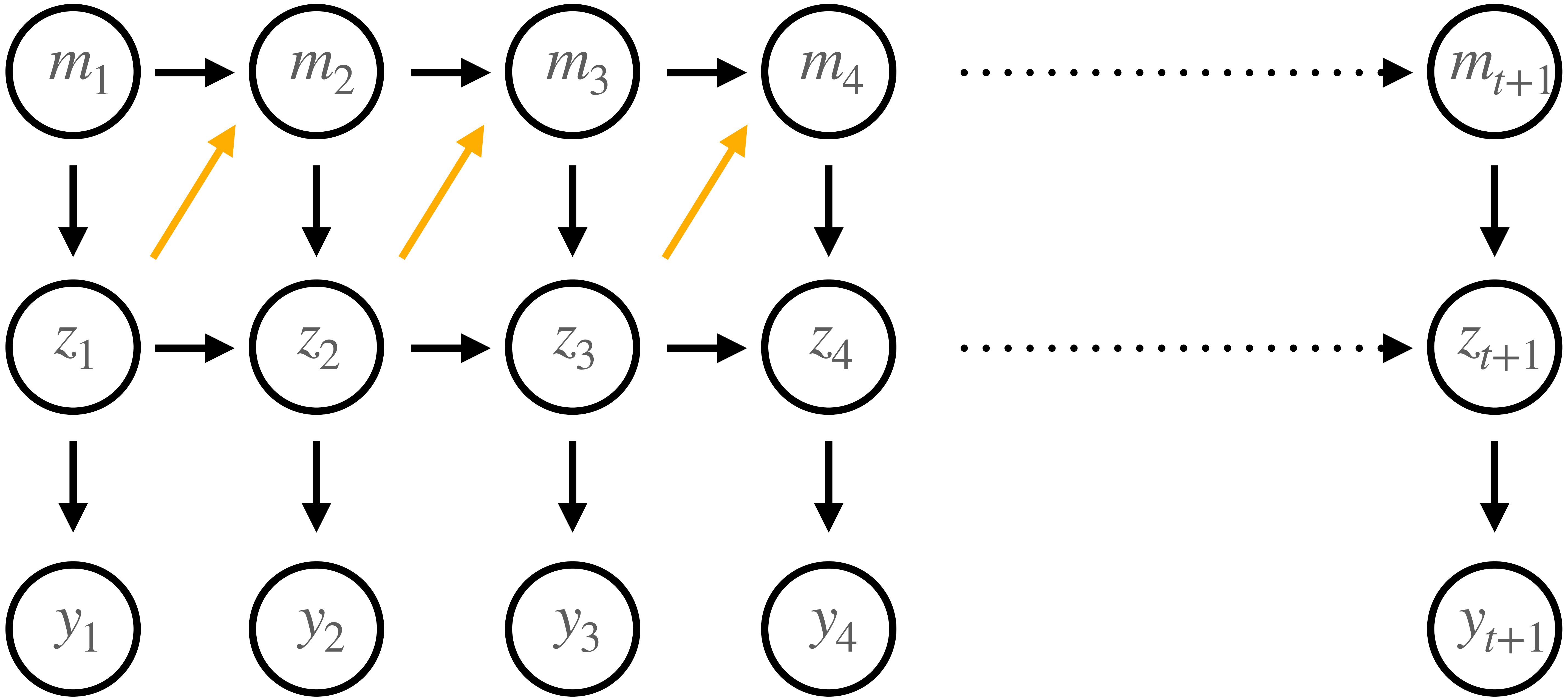
Observations
(Neural Data) $p(\boldsymbol{y}_t | \boldsymbol{z}_t, m_t = k, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t | \mathbf{H}_k \boldsymbol{z}_t + \mathbf{D}_k \boldsymbol{u}_t + \boldsymbol{d}_k, \mathbf{R}_k)$

Recurrent Switching Linear Dynamical Systems

Discrete Latents

Continuous Latents

Observations
(Neural Data)



EM Algorithm and the Evidence Lower Bound (ELBO)

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$\log p(x \mid \theta) = \log \int p(x, z \mid \theta) dz$$ (definition of log-likelihood)

$$= \log \int q(z \mid \phi) \frac{p(x, z \mid \theta)}{q(z \mid \phi)} dz$$ (multiply and divde by q)

$$\geq \int q(z \mid \phi) \log \left(\frac{p(x, z \mid \theta)}{q(z \mid \phi)} \right) dz$$ (apply Jensen)

$$\triangleq F(\phi, \theta)$$ (negative Free Energy)

$$\triangleq \text{ELBO}$$ (evidence lower bound)

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$\log p(x|\theta) = \log \int p(x, z|\theta) dz \qquad \text{(definition of log-likelihood)}$$

$$= \log \int q(z|\phi) \frac{p(x, z|\theta)}{q(z|\phi)} dz \qquad \text{(multiply and divde by } q\text{)}$$

$$\geq \int q(z|\phi) \log \left(\frac{p(x, z|\theta)}{q(z|\phi)} \right) dz \qquad \text{(apply Jensen)}$$

$$\triangleq F(\phi, \theta) \qquad \text{(negative Free Energy)}$$

$$\triangleq \text{ELBO} \qquad \text{(evidence lower bound)}$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$\log p(x|\theta) = \log \int p(x,z|\theta)dz \qquad \text{(definition of log-likelihood)}$$

$$= \log \int q(z|\phi)\frac{p(x,z|\theta)}{q(z|\phi)}dz \qquad \text{(multiply and divde by } q\text{)}$$

$$\geq \int q(z|\phi) \log \left(\frac{p(x,z|\theta)}{q(z|\phi)}\right) dz \qquad \text{(apply Jensen)}$$

$$\triangleq F(\phi,\theta) \qquad \text{(negative Free Energy)}$$

$$\triangleq \text{ELBO} \qquad \text{(evidence lower bound)}$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$\log p(x|\theta) = \log \int p(x,z|\theta)dz$$ (definition of log-likelihood)

$$= \log \int q(z|\phi)\frac{p(x,z|\theta)}{q(z|\phi)}dz$$ (multiply and divde by q)

$$\geq \int q(z|\phi) \log \left(\frac{p(x,z|\theta)}{q(z|\phi)}\right) dz$$ (apply Jensen)

$$\triangleq F(\phi,\theta)$$ (negative Free Energy)

$$\triangleq \text{ELBO}$$ (evidence lower bound)

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz$$

$$= \int q(z|\phi) \log \left[\frac{p(x|\theta) p(z|x, \theta)}{q(z|\phi)} \right] dz$$

$$= \int q(z|\phi) \log p(x|\theta) + \int q(z|\phi) \log \left[\frac{p(z|x, \theta)}{q(z|\phi)} \right] dz$$

$$= \log p(x|\theta) - KL\Big(q(z|\phi) || p(z|x, \theta) \Big)$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$
\begin{aligned}
F(\phi, \theta) &= \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log \left[\frac{p(x|\theta) p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log p(x|\theta) + \int q(z|\phi) \log \left[\frac{p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \log p(x|\theta) - KL\Big(q(z|\phi) || p(z|x, \theta) \Big)
\end{aligned}
$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$
\begin{aligned}
F(\phi, \theta) &= \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log \left[\frac{p(x|\theta) p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log p(x|\theta) + \int q(z|\phi) \log \left[\frac{p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \log p(x|\theta) - KL\Big(q(z|\phi) || p(z|x, \theta)\Big)
\end{aligned}
$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$
\begin{aligned}
F(\phi, \theta) &= \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log \left[\frac{p(x|\theta) p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \int q(z|\phi) \log p(x|\theta) + \int q(z|\phi) \log \left[\frac{p(z|x, \theta)}{q(z|\phi)} \right] dz \\
&= \log p(x|\theta) - KL\Big(q(z|\phi) || p(z|x, \theta) \Big)
\end{aligned}
$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \log p(x|\theta) - KL\Big(q(z|\phi)||p(z|x, \theta)\Big)$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \log p(x|\theta) - KL\Big(q(z|\phi)||p(z|x, \theta)\Big)$$

- **KL Divergence**
  - Measures how different two distributions are

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \log p(x|\theta) - KL\Big(q(z|\phi)||p(z|x, \theta)\Big)$$

- **KL Divergence**
  - Measures how different two distributions are
  - Is greater than or equal to 0
    - $F(\phi, \theta)$ is a lower bound on the evidence, $\log p(x|\theta)$
    - When $q(z|\phi) = p(z|x, \theta)$, the bound is tight.

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \log p(x|\theta) - KL\Big(q(z|\phi)||p(z|x, \theta)\Big)$$

- **KL Divergence**
  - Measures how different two distributions are
  - Is greater than or equal to 0
    - $F(\phi, \theta)$ is a lower bound on the evidence, $\log p(x|\theta)$
    - When $q(z|\phi) = p(z|x, \theta)$, the bound is tight.

- **E-step**: Update ϕ by setting $q(z|\phi) = p(z|x, \theta)$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz$$
$$= \int q(z|\phi) \log p(x, z|\theta) dz - \int q(z|\phi) \log q(z|\phi) dz.$$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

EM Derivation: The ELBO

$$F(\phi, \theta) = \int q(z|\phi) \log \left[\frac{p(x, z|\theta)}{q(z|\phi)} \right] dz$$
$$= \int q(z|\phi) \log p(x, z|\theta) dz - \int q(z|\phi) \log q(z|\phi) dz.$$

- **M-step**: Update θ by maximizing the expected total data likelihood, $\int q(z|\phi) \log p(x, z|\theta) dz$

https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf

Resources

- Probabilistic Machine Learning Book 2, Chap. 29
  - https://probml.github.io/pml-book/book2.html
- Intro to LVM Notes from Princeton Course
  - https://pillowlab.princeton.edu/teaching/statneuro2020/notes/notes18_LatentVariableModels.pdf