Model's detailed specification for black box model

① sample $z' \sim P_\theta(z|g) = \prod_j q_{\phi_G}^{(j)}(z_G^{(j)}|g)$

↳ the prior

② sample $x \sim P_\theta(Y|z,g,u) = P(Y|M,\Sigma)$

$Y: M \times T$ (4×100 for case study)

\# species \#time steps

$Y = [y_1, y_2 \cdots y_T]$

$M = \psi(X)$

$\Sigma = \rho(X, z)$

$\dot{x} = w_1^+(x,\psi) - x \odot w_2^+(x,\psi)$ $X = [x_1, x_2 \cdots x_T]$

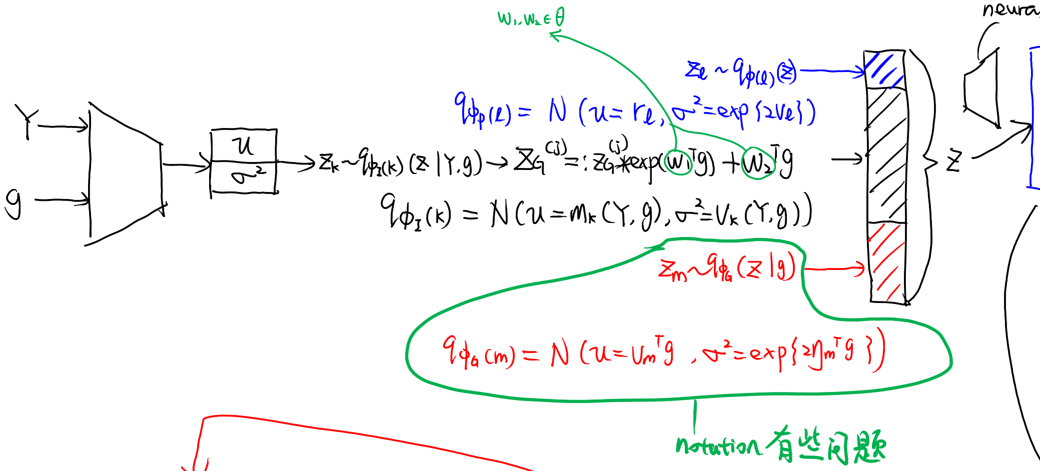
$\dot{v} = w_3^+(v,x,\psi) - v \odot w_4^+(v,x,\psi)$ $V = [v_1, v_2, \cdots v_T]$

θ in general : parameters that can be adjusted during learning

ODE function $\dot{x} = w(x, \psi; \theta)$

neural network weights

neural network $\psi = \{z_P, z_G, z_I, u, g\}$



$\dot{x} = w_1^+(x,\psi) - x \odot w_2^+(x,\psi)$

x_0

$\dot{v} = w_3^+(v,x,\psi) - v \odot w_4^+(v,x,\psi)$

v_0

ODE function for black-box model

w^+ : a neural network with softplus activation function

$X = [x_1, x_2, x_3, \cdots x_T]$

$\Sigma = [v_1, v_2, v_3 \cdots v_T]$

$M = X$ (if the state space $\vec{x}$ = observation space $\vec{y}$)

otherwise $M = \psi(X)$ showcase an observation process

$Y \sim P(Y|M,\Sigma)$

ELBO of $\log P(Y|g,u)$

$E_{q_\phi(z|Y,g,u)}[\log P_\theta(Y|z,g,u) + \log P_\theta(z|g) - \log q_\phi(z|Y,g)]$

↳ closed form ?

iwae → a tweak at cost function to achieve a tighter lower bound

DReG → less variance for gradient estimation

$= \log P_\theta(Y|z,g,u) - KL(q_\phi(z|Y,g) \| P_\theta(z|g))$

batch size = N

$$\begin{bmatrix} ELBO_1 \rightarrow \log P(Y_1|g_1,u_1) \\ ELBO_2 \rightarrow \log P(Y_2|g_2,u_2) \\ \vdots \\ ELBO_N \rightarrow \log P(Y_N|g_N,u_N) \end{bmatrix} \rightarrow \frac{1}{N}\mathrm{Sum}([\,]) = loss$$