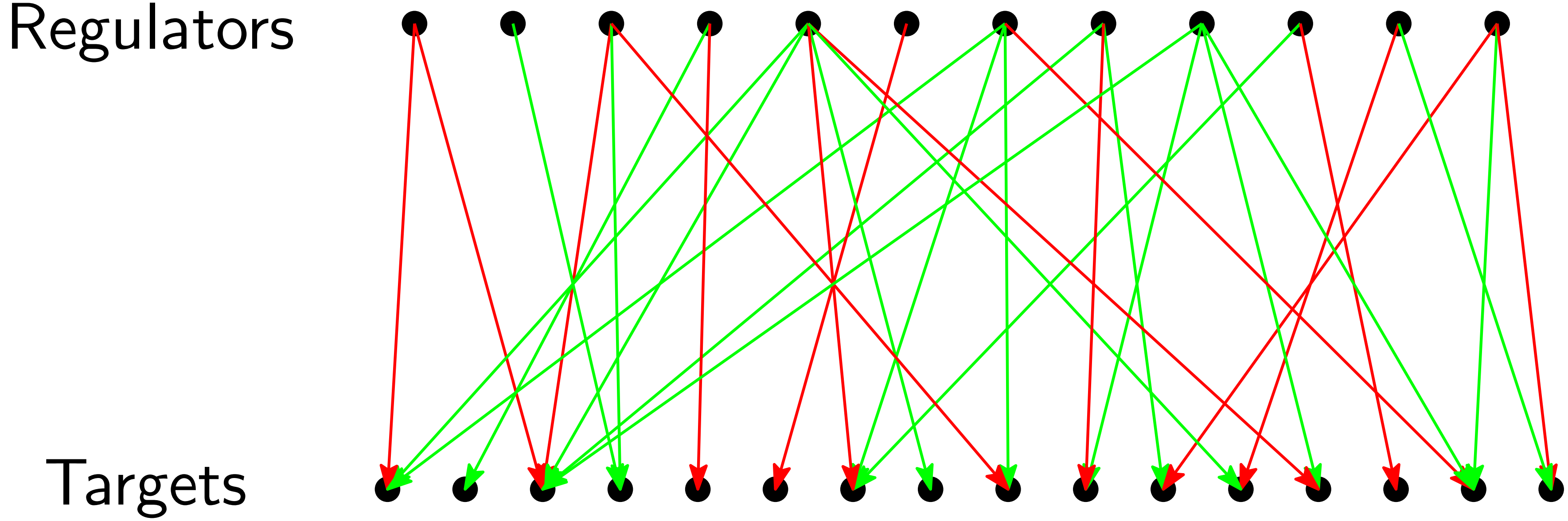
Detecting Deregulated Genes

September 30, 2015

With Étienne Birmelé, Julien Chiquet, Mohamed Elati, Pierre Neuvial, Rémy Nicolle, François Radvanyi

Thomas Picchetti

Problem

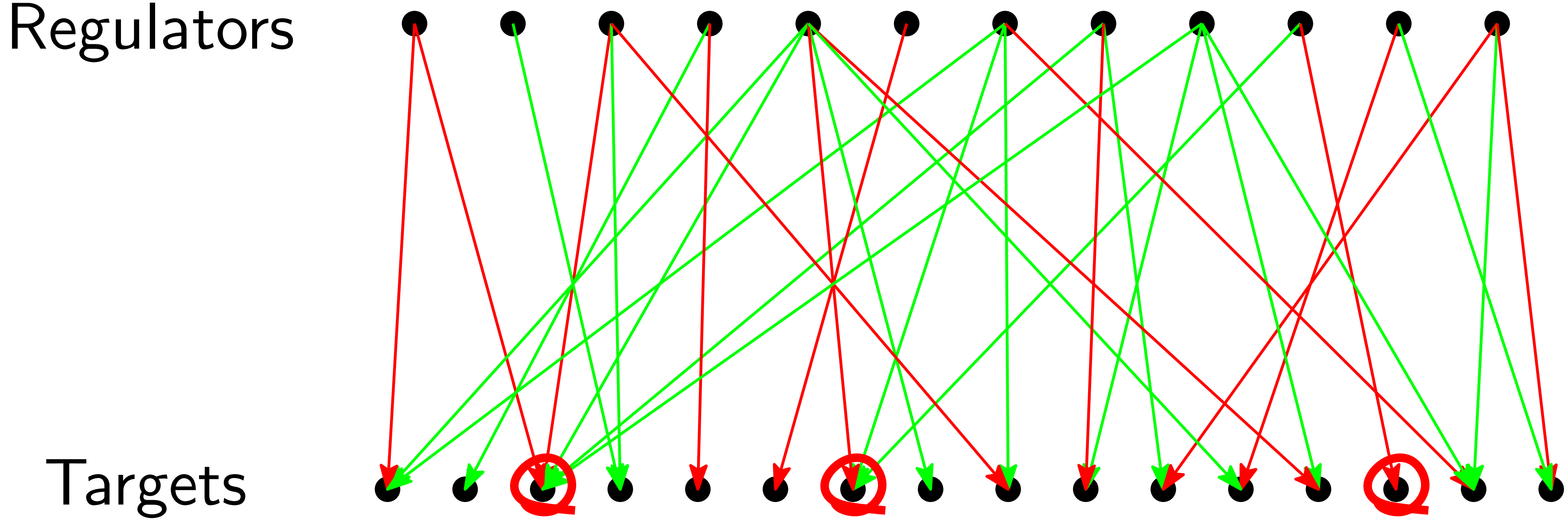


Input :

- a gene regulation network. Target genes have activators and inhibitors among regulator genes.
- a set of expression profiles (e.g. RNA microarrays)

Question: In each profile, which target genes disobey their regulators ?

Problem

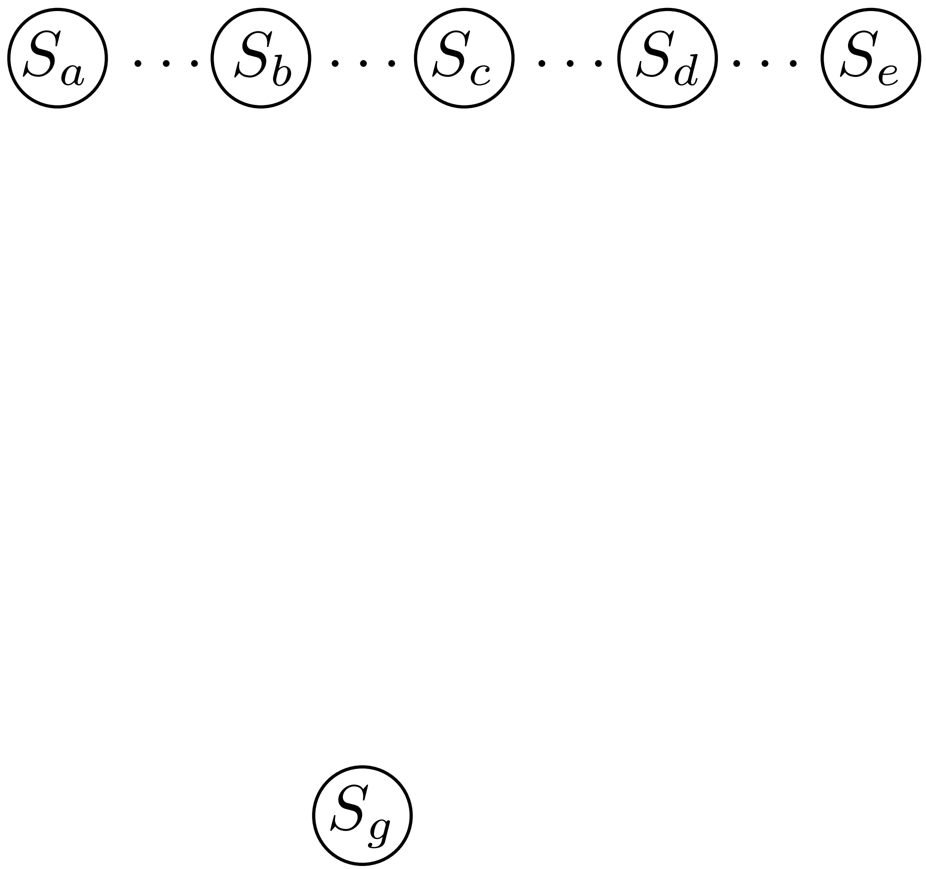


Input :

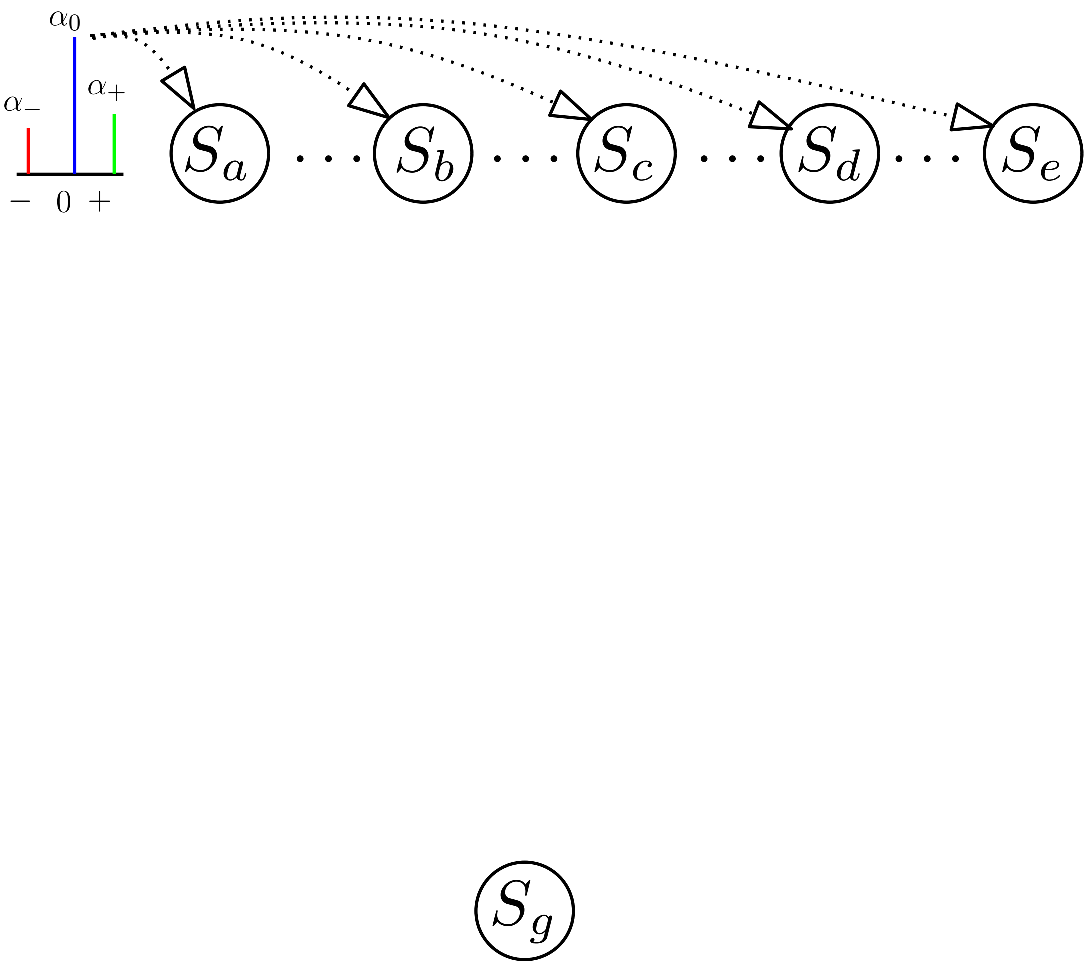
- a gene regulation network. Target genes have activators and inhibitors among regulator genes.
- a set of expression profiles (e.g. RNA microarrays)

Question: In each profile, which target genes disobey their regulators ?

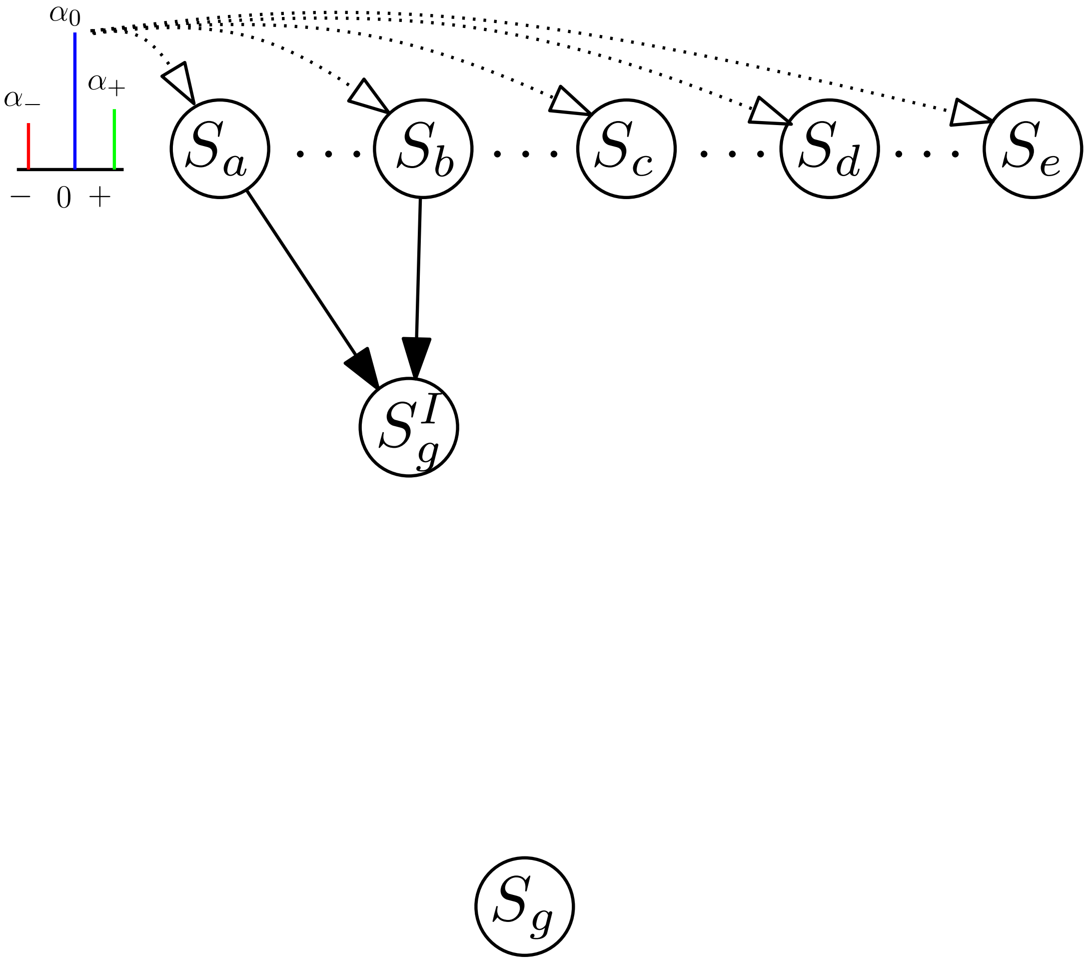
A model that allows deregulation



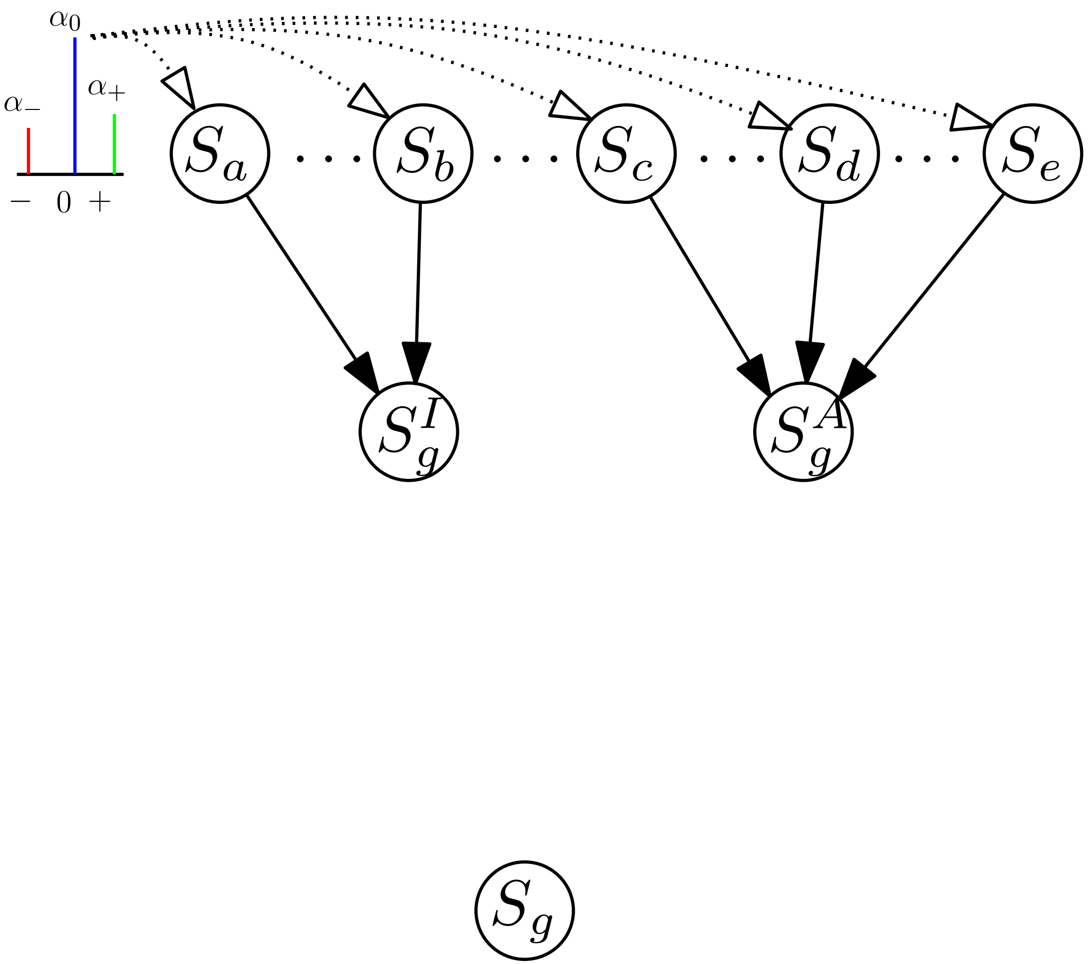
A model that allows deregulation



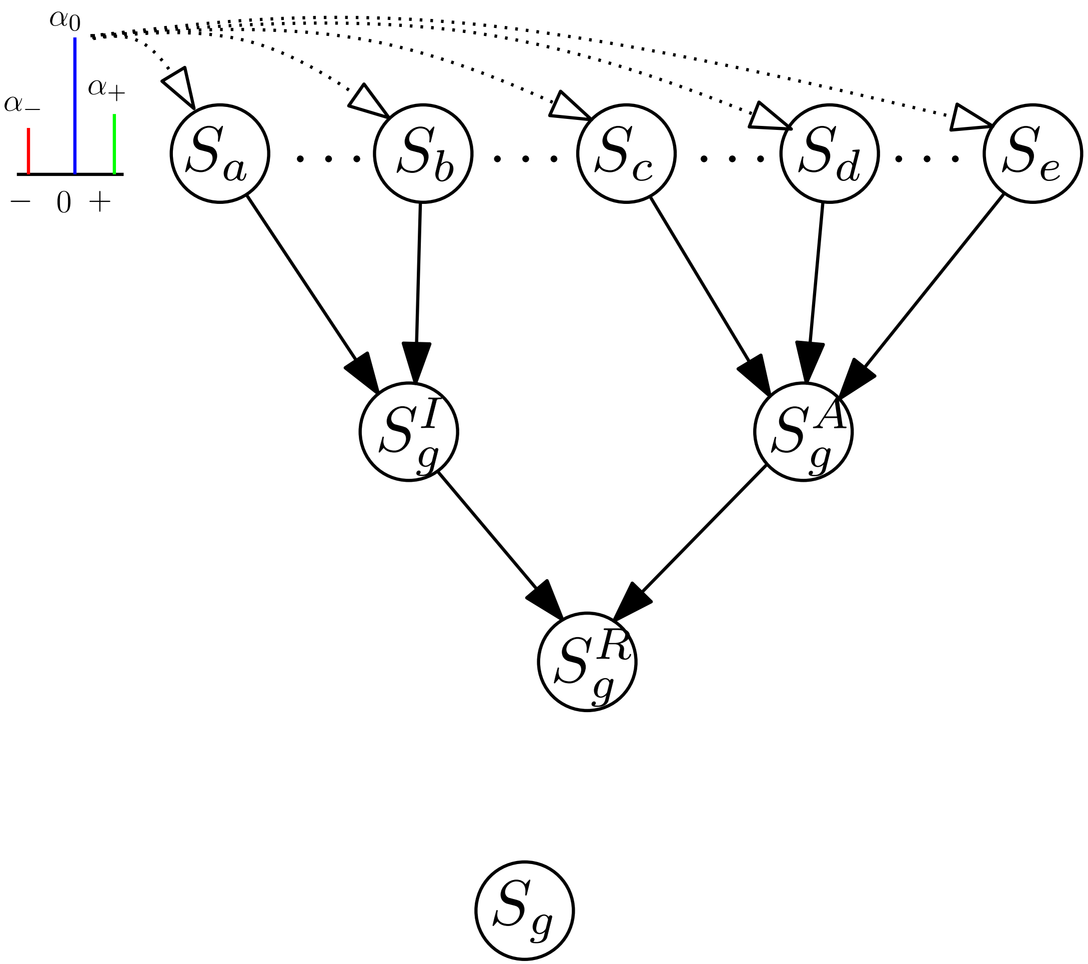
A model that allows deregulation



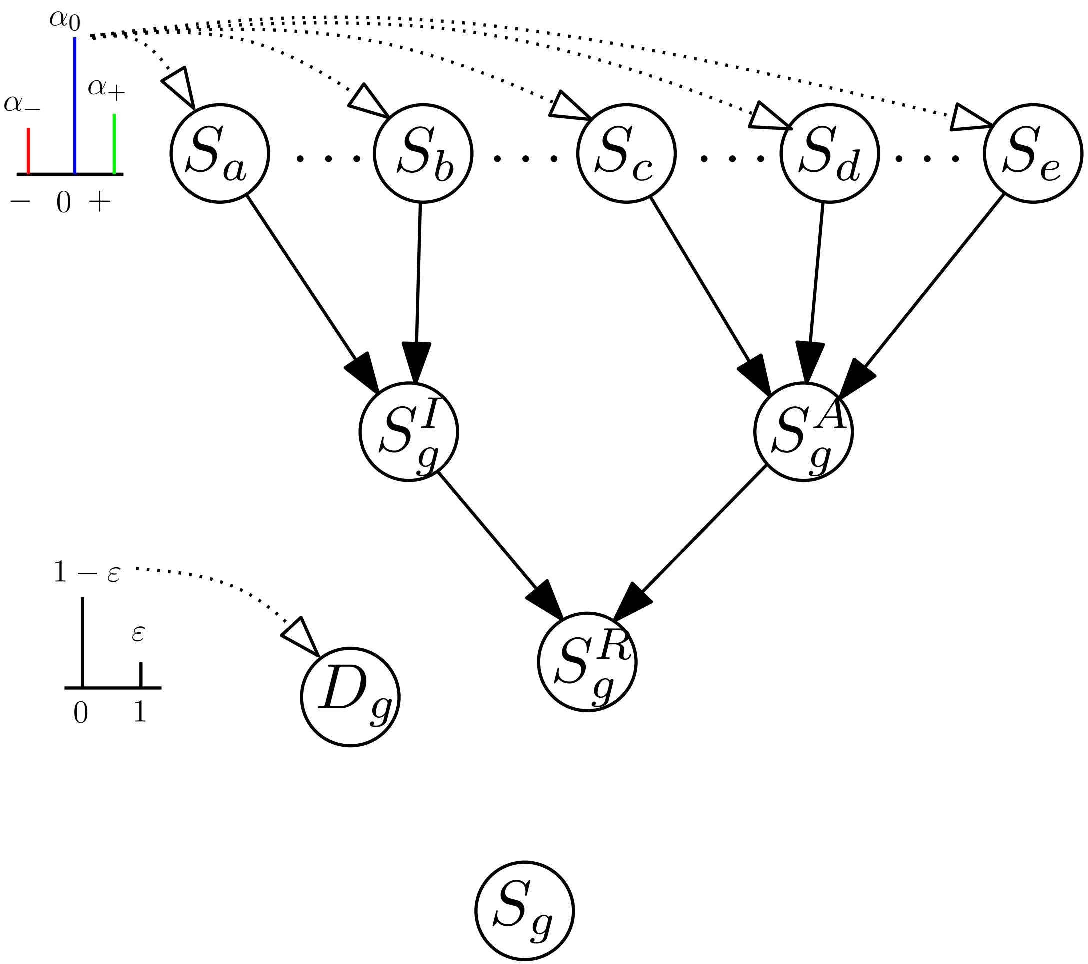
A model that allows deregulation



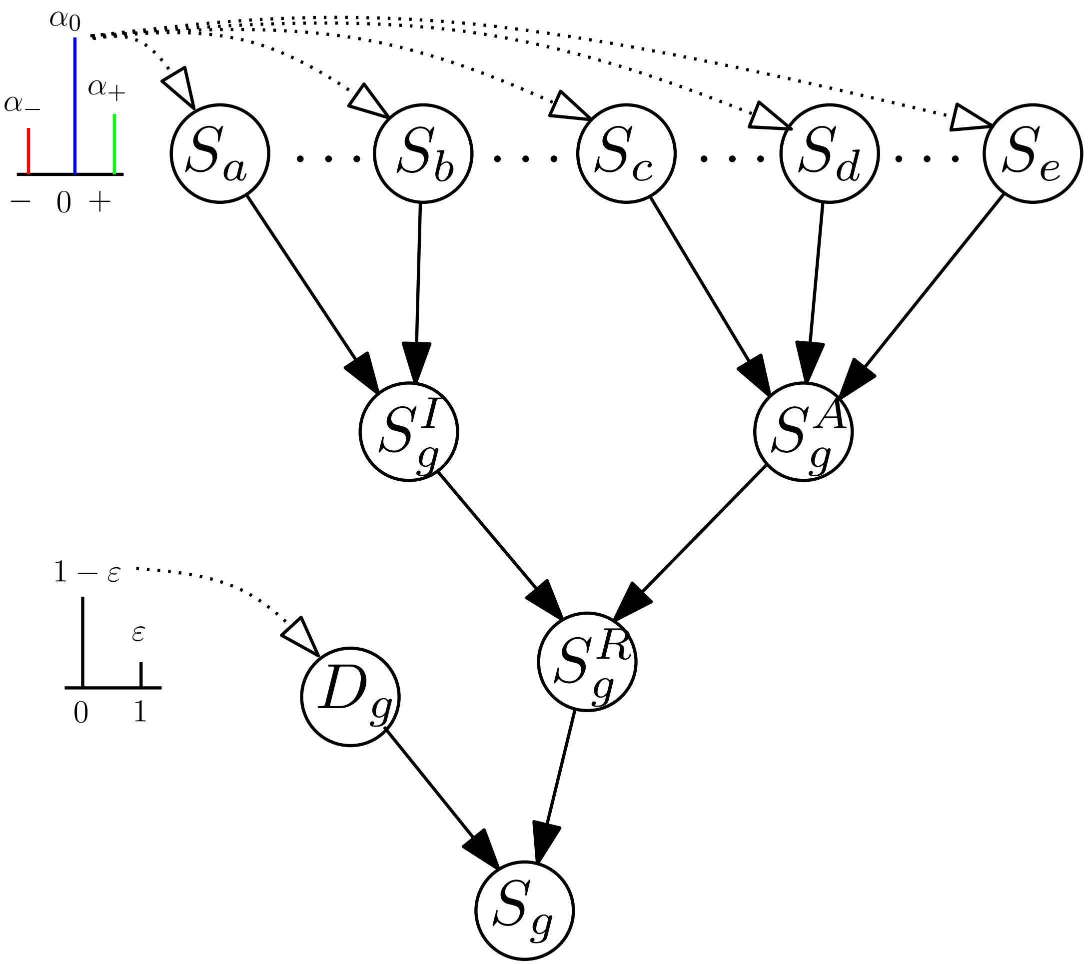
A model that allows deregulation



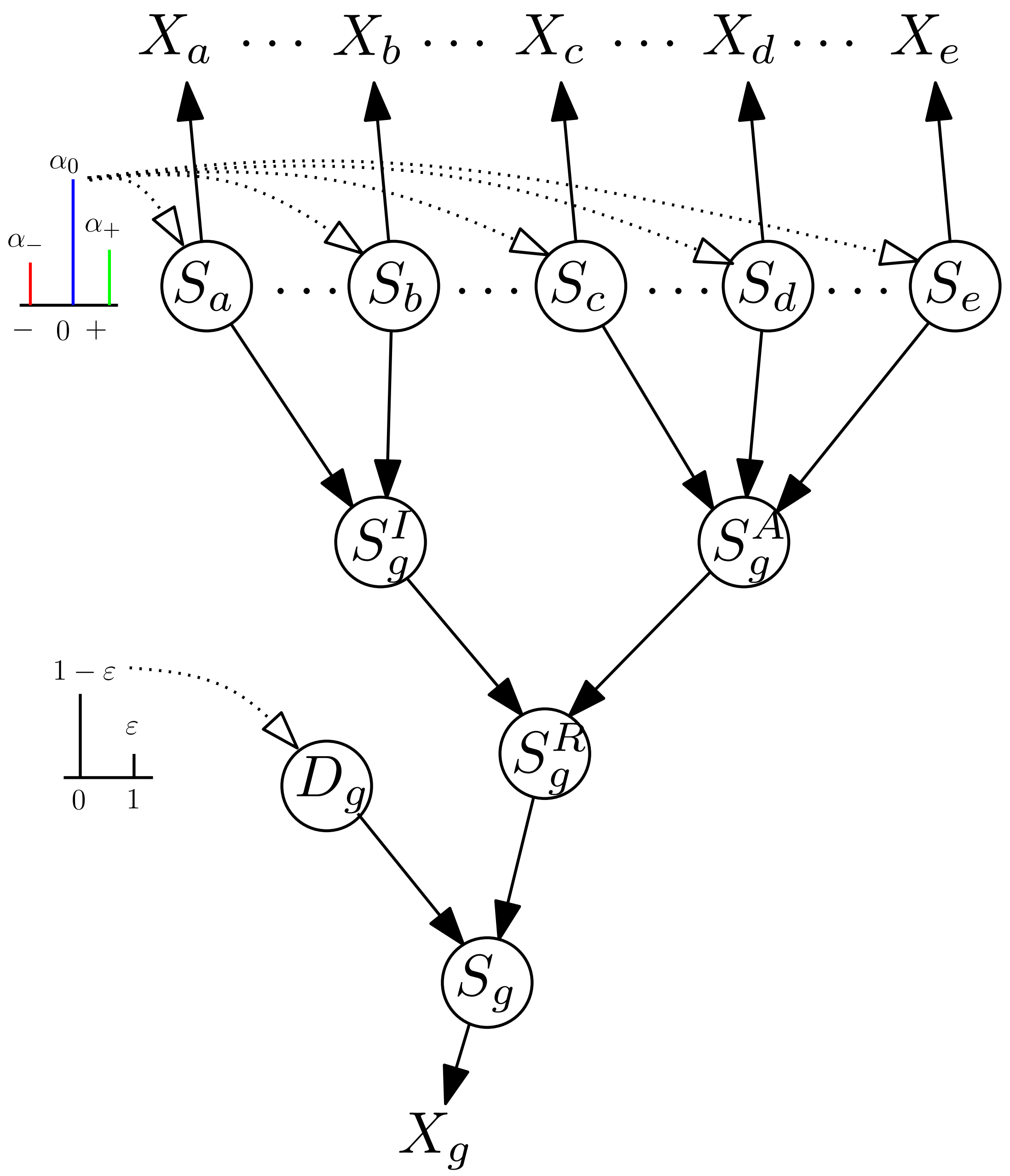
A model that allows deregulation



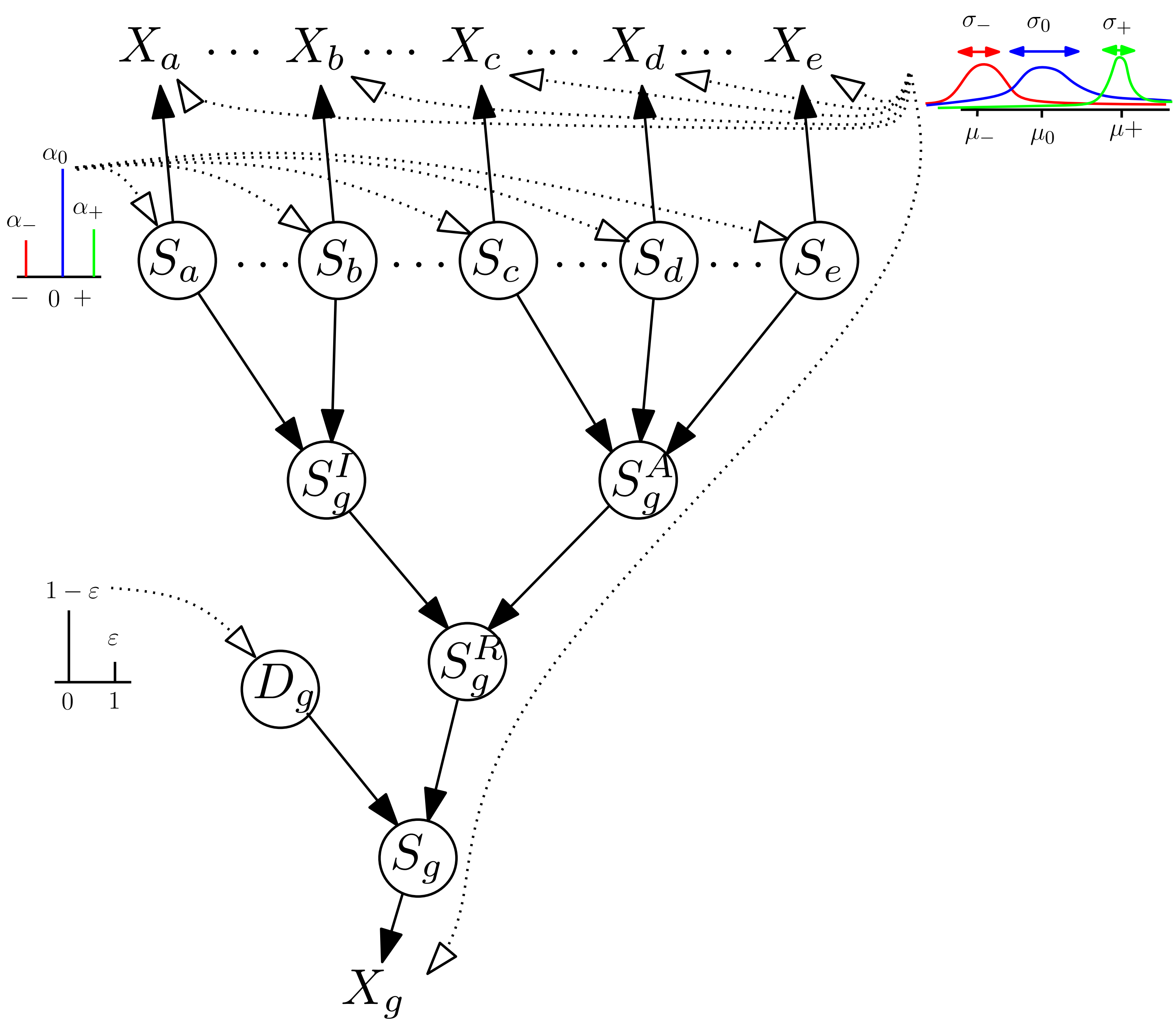
A model that allows deregulation



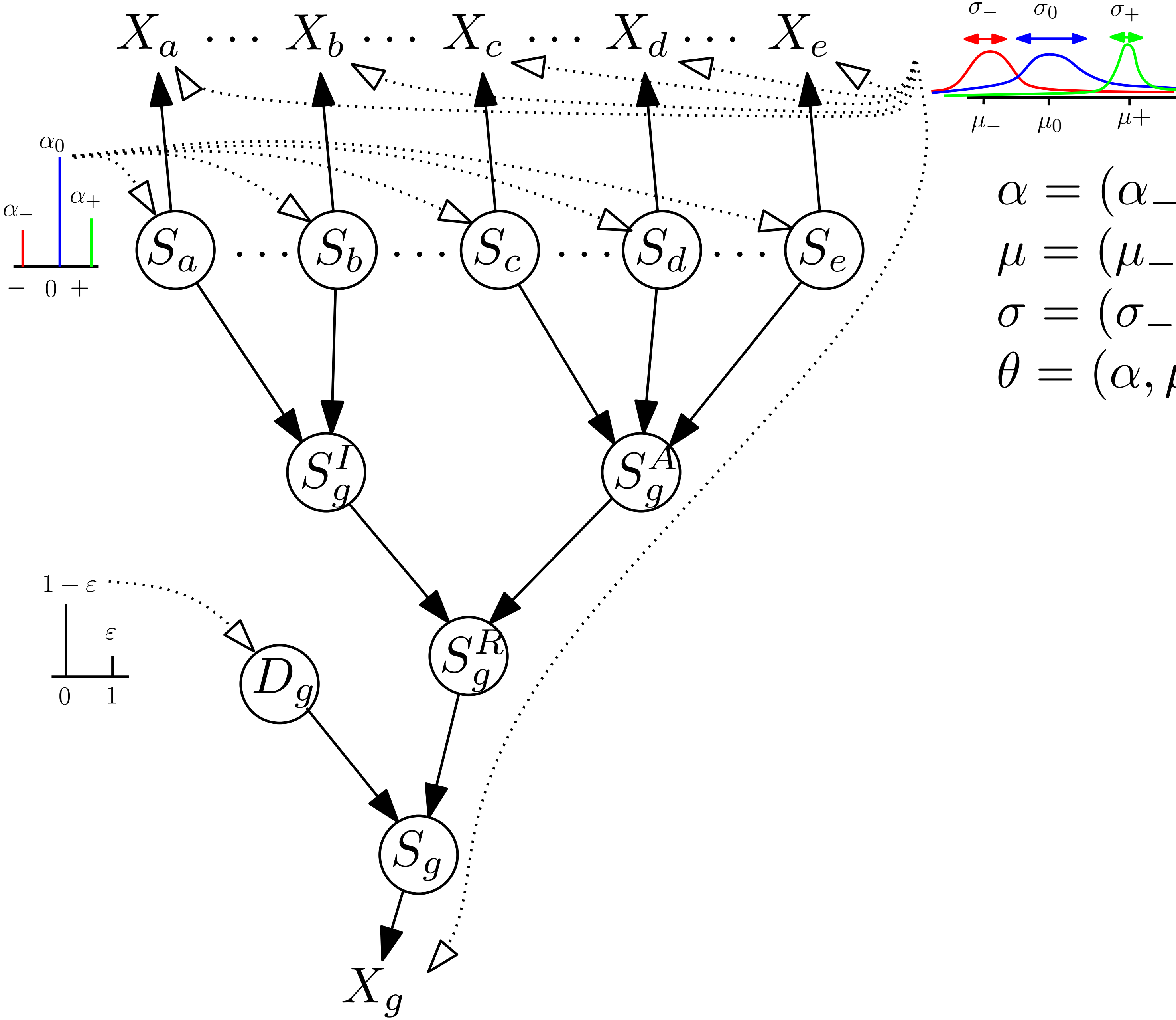
A model that allows deregulation



A model that allows deregulation

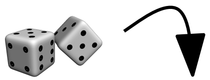


A model that allows deregulation



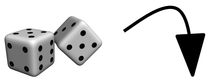
$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

A probabilistic model



Regulator states $\in \{-, 0, +\}$

A probabilistic model



Regulator states $\in \{-, 0, +\}$

Regulation

Target states $\in \{-, 0, +\}$

A probabilistic model

Regulator states $\in \{-, 0, +\}$

 $+$ 

Regulation   Deregulation

Target states $\in \{-, 0, +\}$

A probabilistic model

Regulator states $\in \{-, 0, +\}$

 $+$ 

Regulation   Deregulation

Target states $\in \{-, 0, +\}$

Hidden variables : Z

A probabilistic model

Regulator states $\in \{-, 0, +\}$

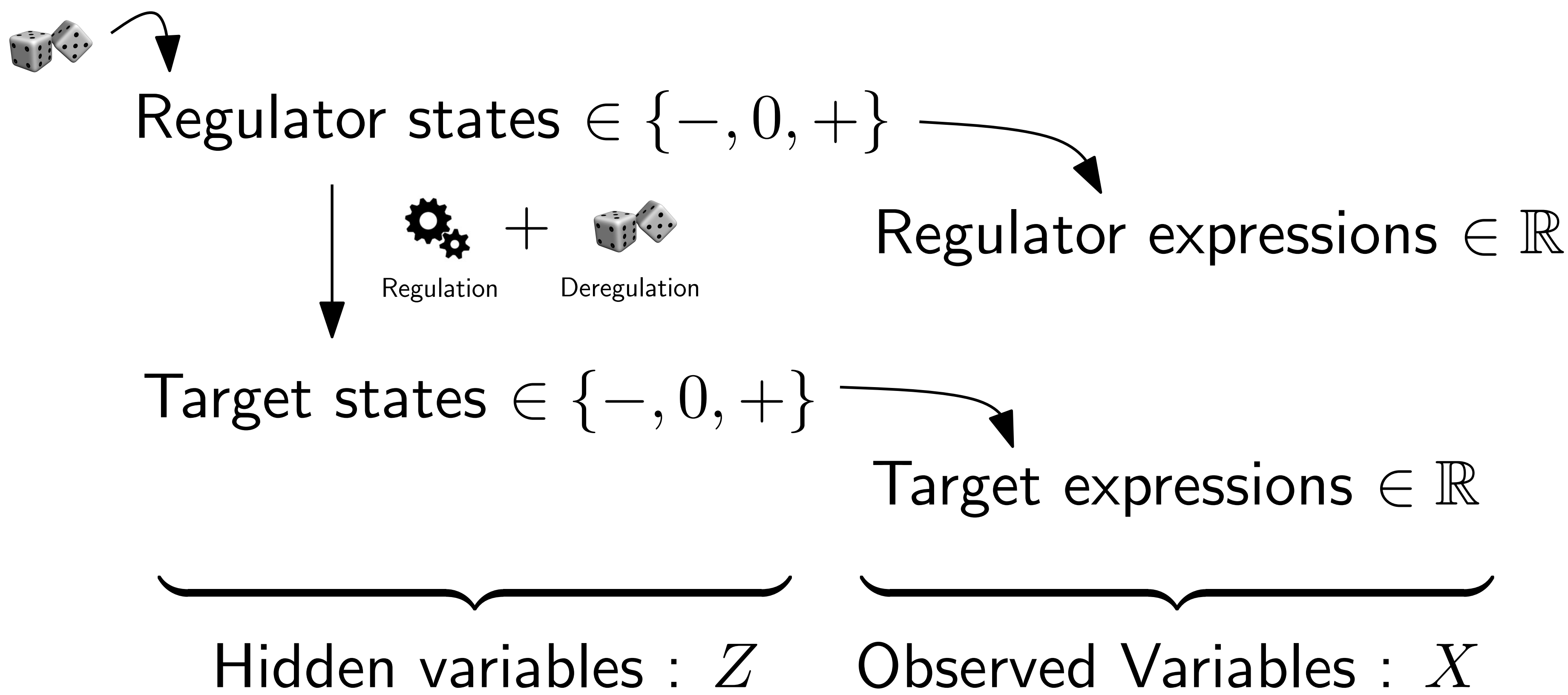
 + 

Regulation   Deregulation

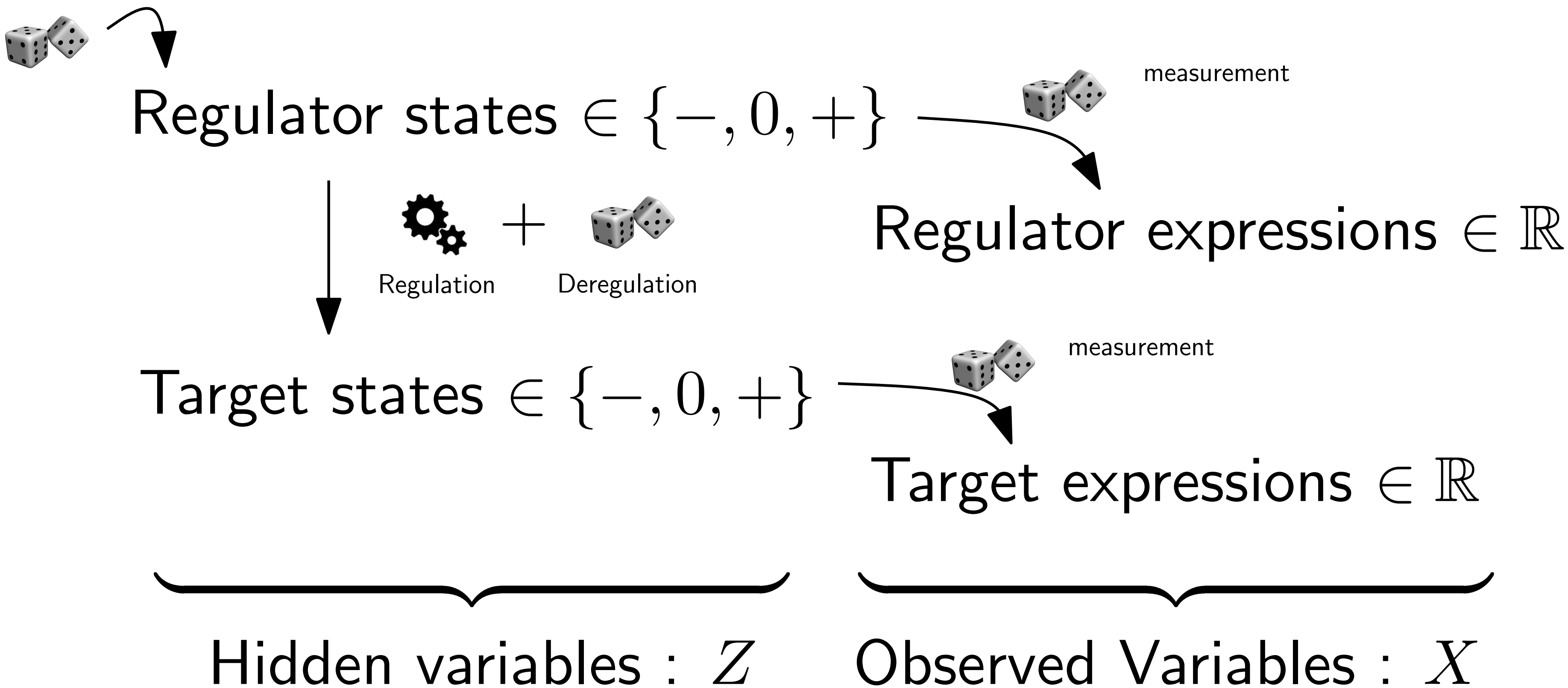
Target states $\in \{-, 0, +\}$

Hidden variables : Z

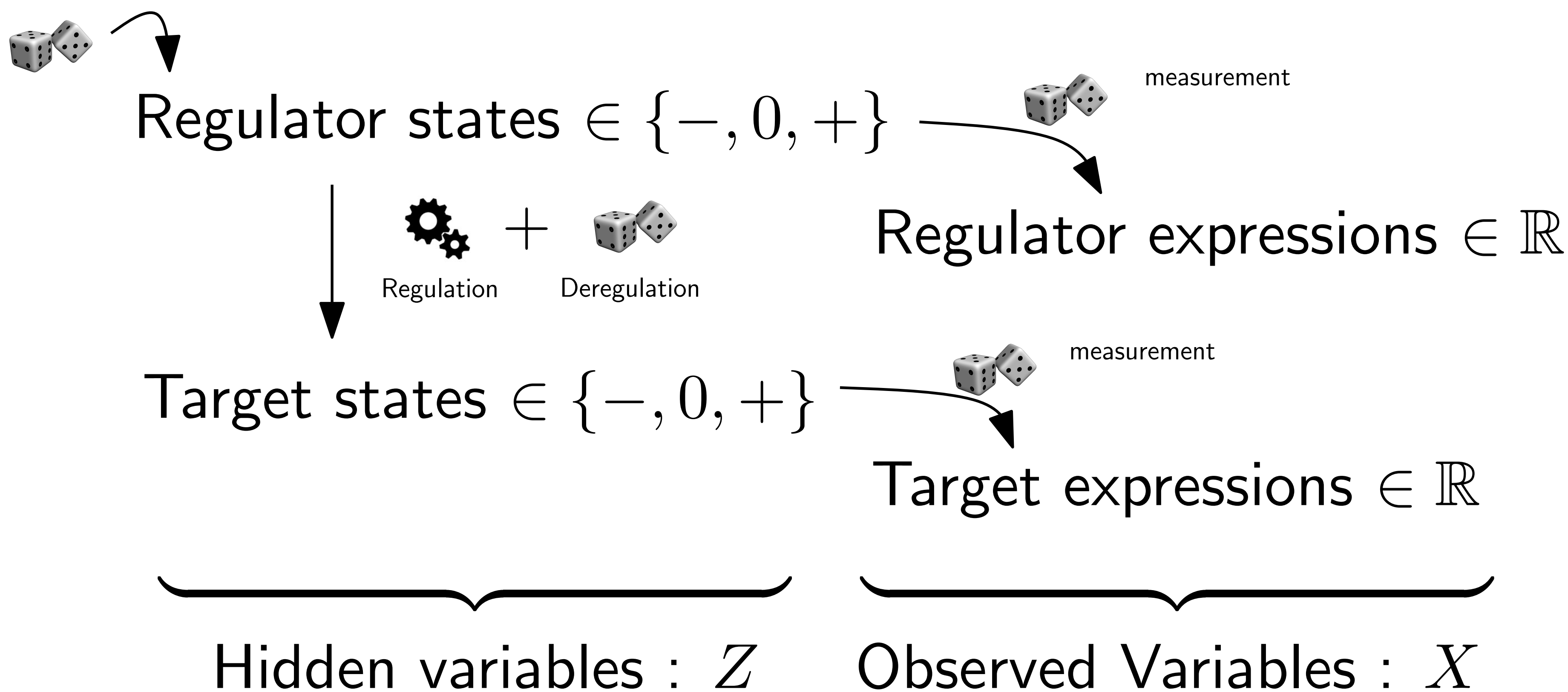
A probabilistic model



A probabilistic model

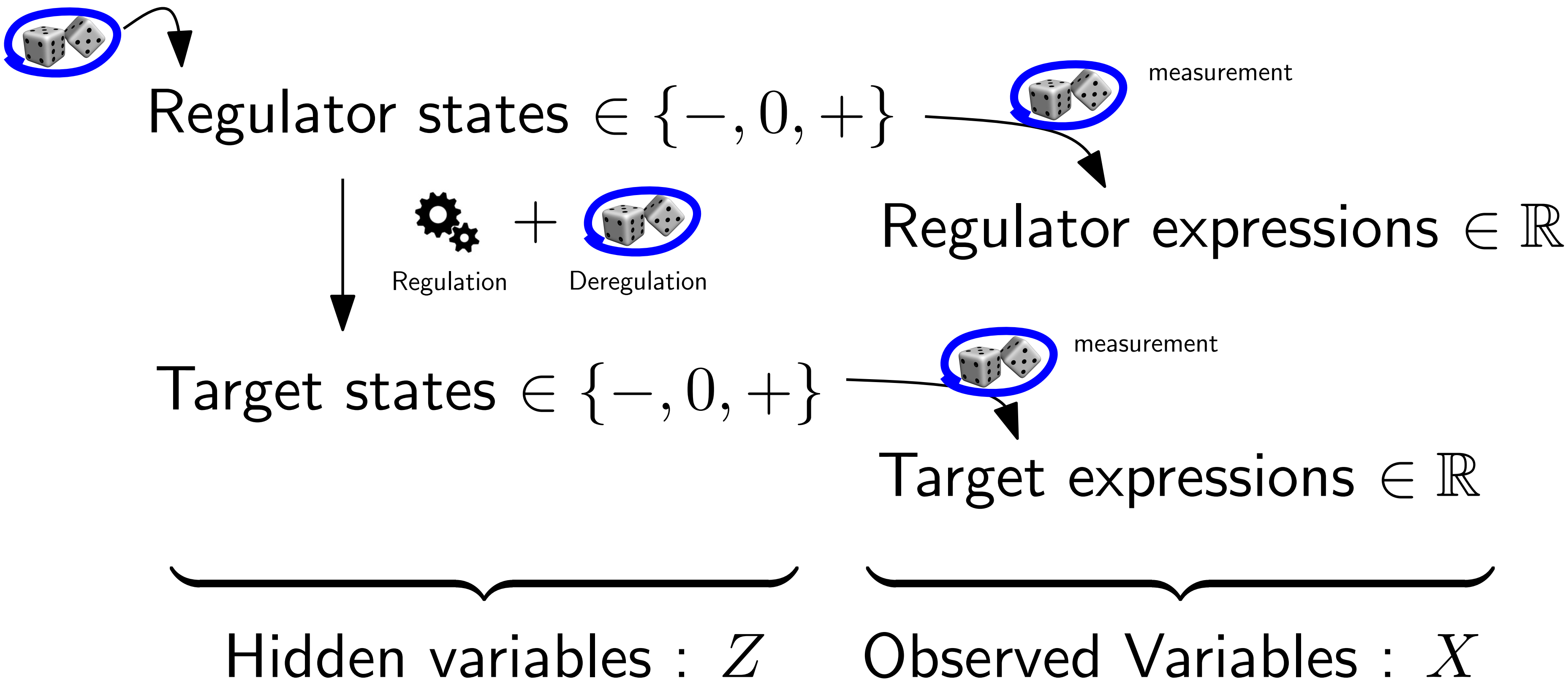


A probabilistic model



Use $P(X, Z)$ to compute $P(Z|X)$ and even $P(\text{gene } \mathbf{g} \text{ deregulated}|X)$

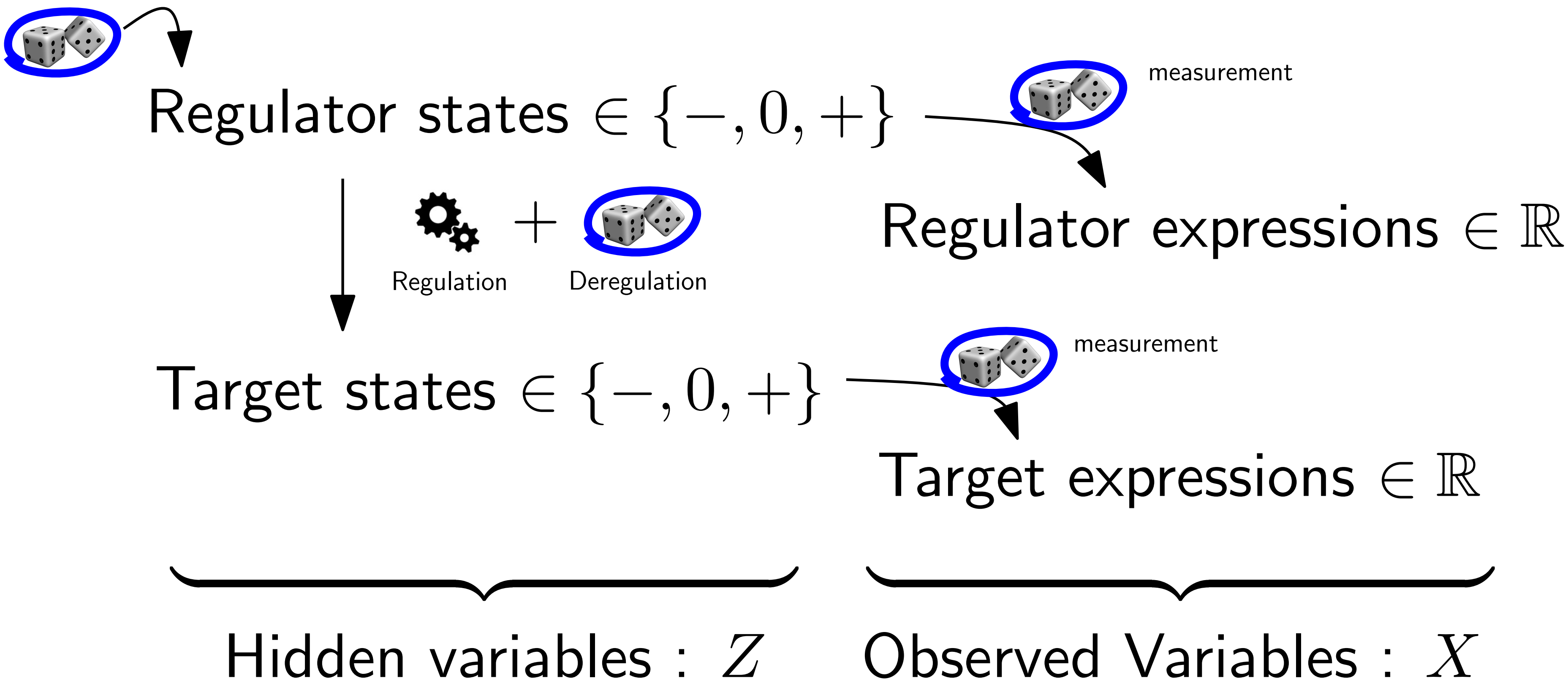
A probabilistic model



Use $P(X, Z)$ to compute $P(Z|X)$ and even $P(\text{gene } \mathbf{g} \text{ deregulated}|X)$

$P(X, Z)$ depends on unknown parameters.

A probabilistic model



Use $P(X, Z)$ to compute $P(Z|X)$ and even $P(\text{gene } \mathbf{g} \text{ deregulated}|X)$

$P(X, Z)$ depends on unknown parameters.

$\rightarrow$ To be inferred from the data.

EM...

We want parameters θ to maximize the likelihood $\mathcal{L} = P(X|\theta)$
Given θ_0, two steps provide a better θ_1:

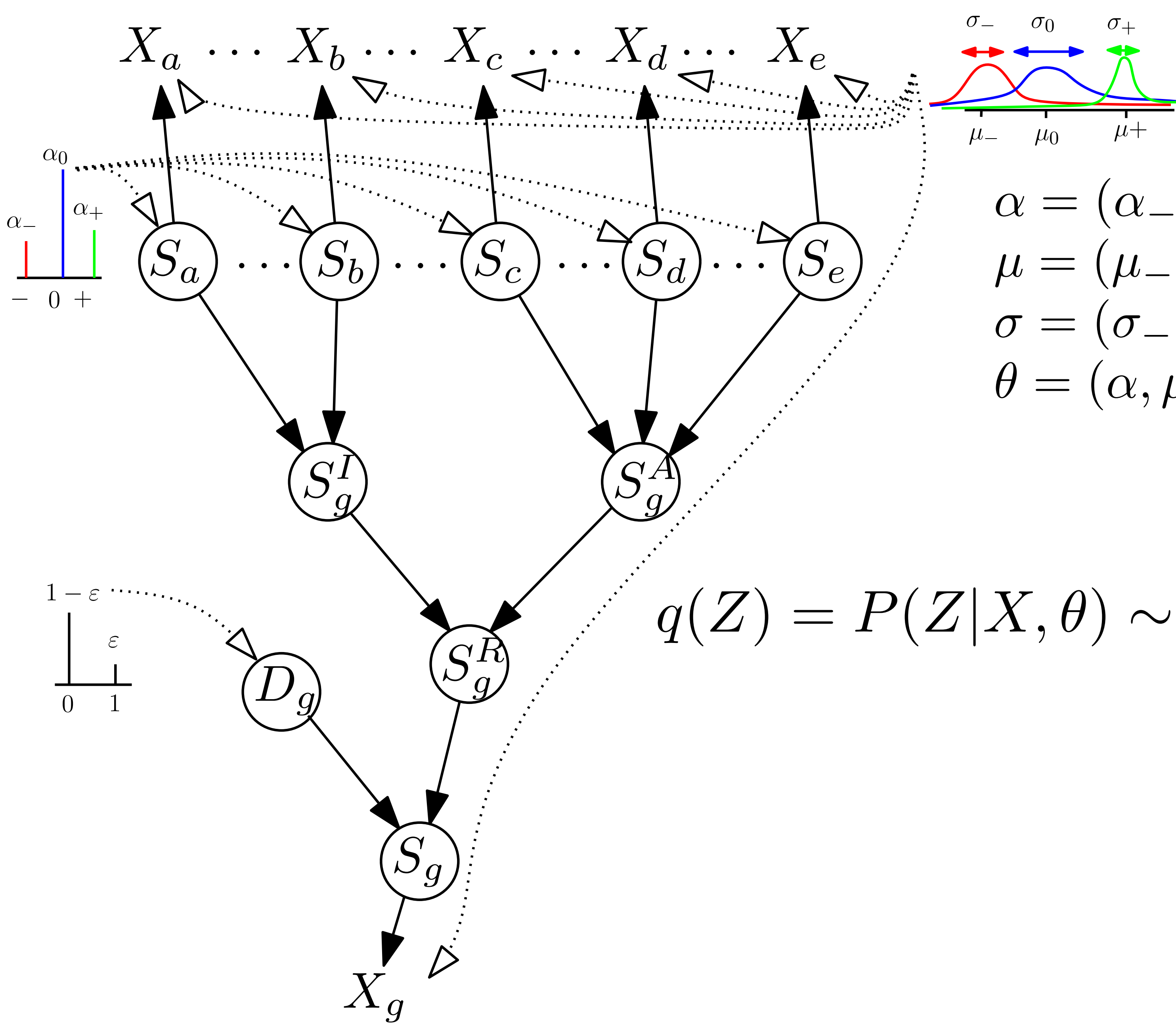
- Step E: compute $q(Z) = P(Z|X, \theta_0)$, the posterior distribution of hidden variables
- Step M: q fixed, the expectation (under q) of $\log P(X, Z|\theta)$ is a function of θ:

$$f(\theta) = \sum_Z q(Z) \log P(X, Z|\theta)$$

$\theta_1 = \operatorname{argmax} f$ is a better parameter set.

In our model, Step M is easy, Step E uses Belief Propagation (detailed later)

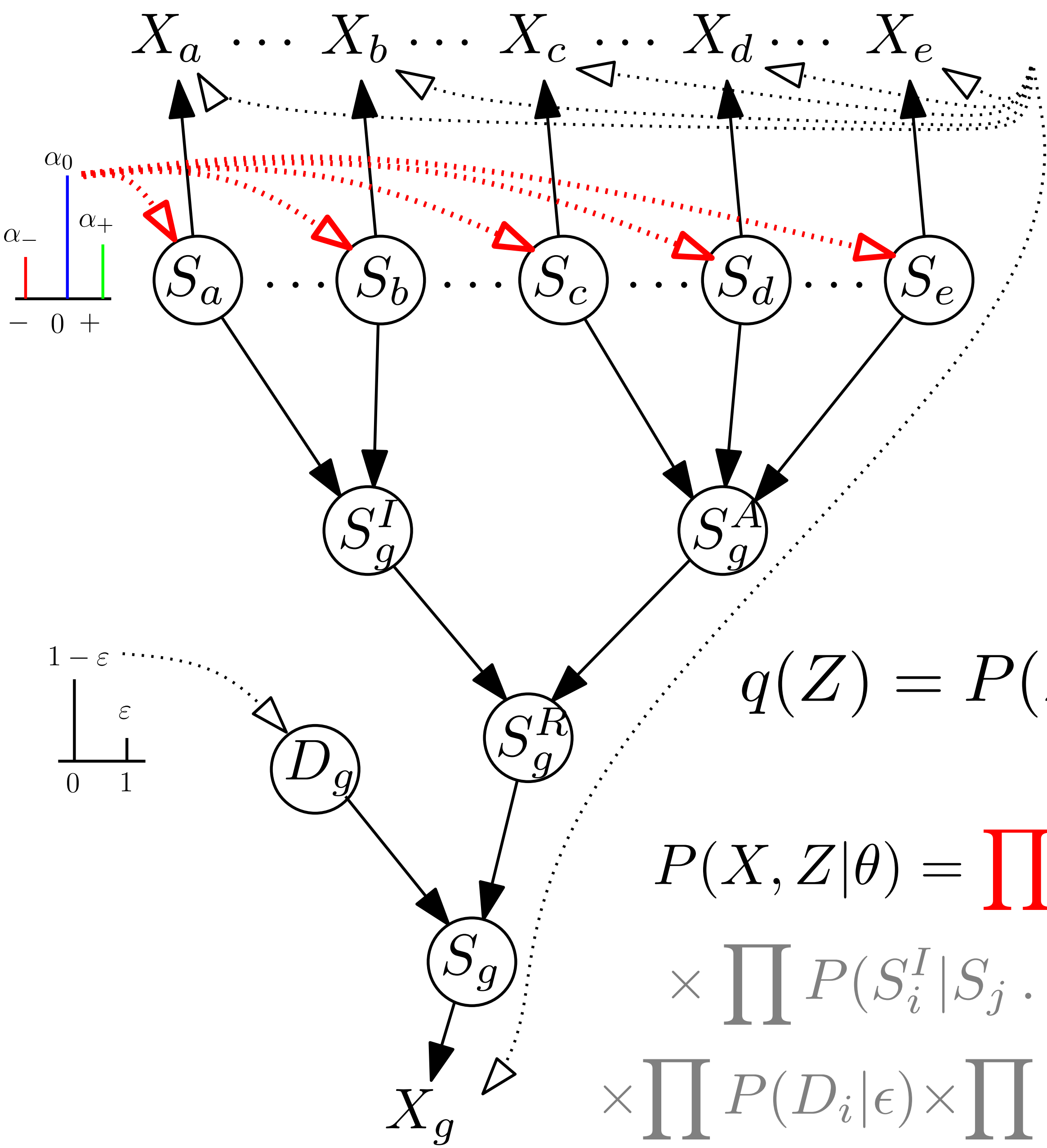
Detailed model : A Factor Graph

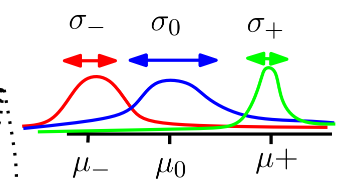


$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X, \theta) \sim P(Z, X|\theta)$$

Detailed model : A Factor Graph



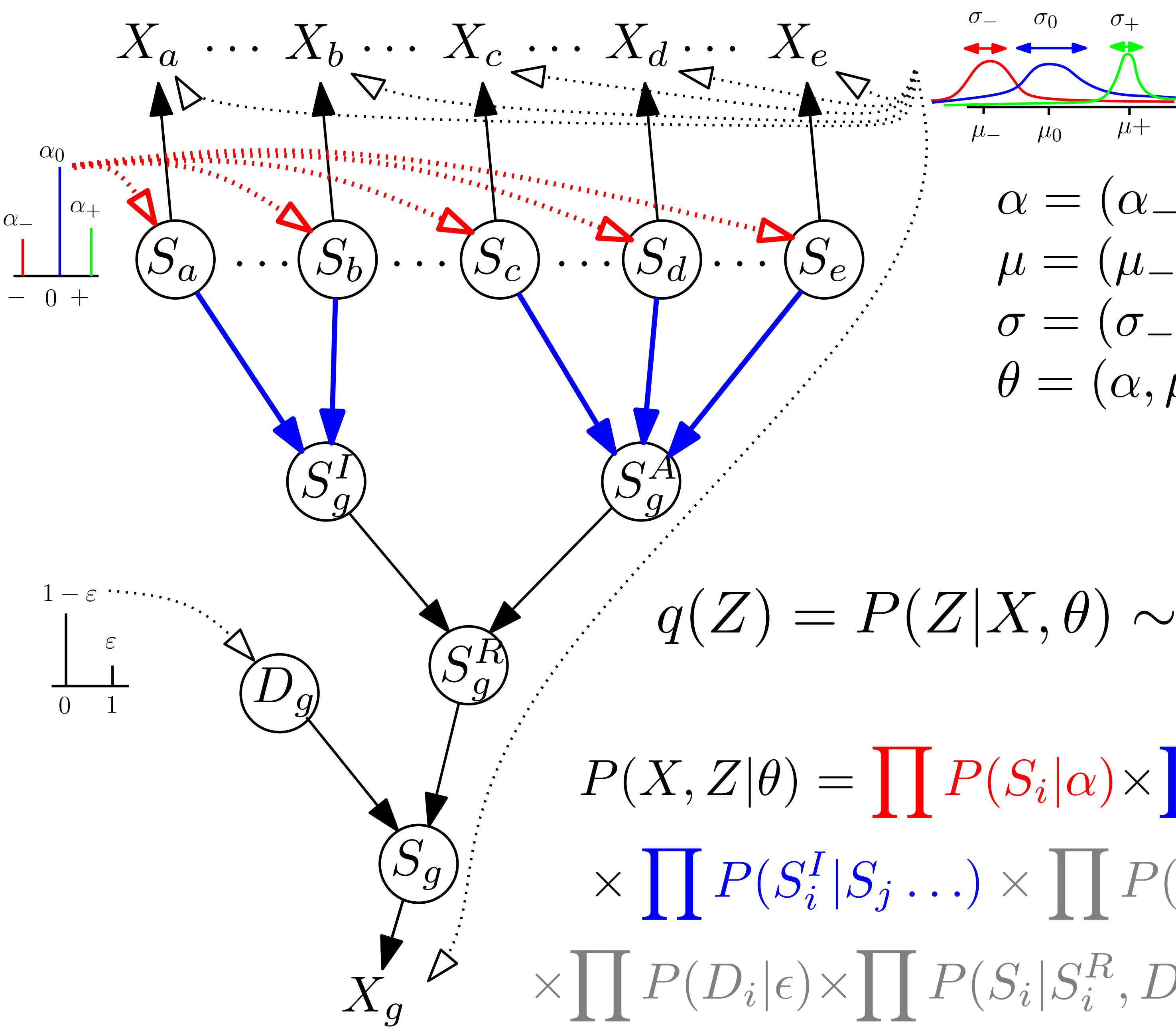


$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X,\theta) \sim P(Z,X|\theta)$$

$$P(X,Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots)$$
$$\times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A)$$
$$\times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

Detailed model : A Factor Graph

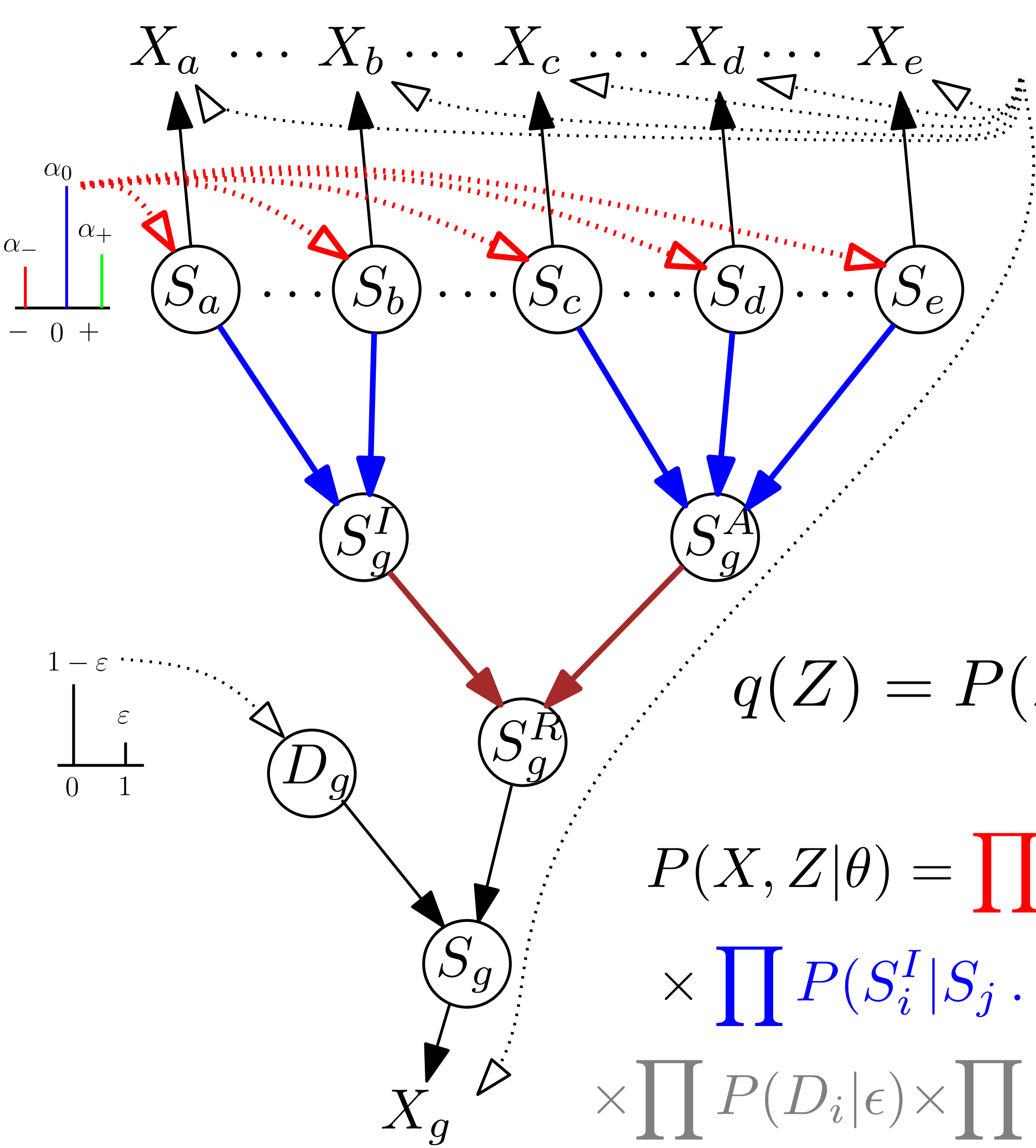


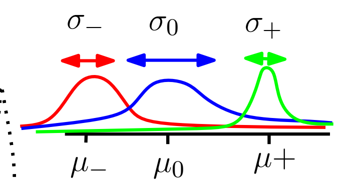
$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X, \theta) \sim P(Z, X|\theta)$$

$$P(X, Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots)$$
$$\times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A)$$
$$\times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

Detailed model : A Factor Graph



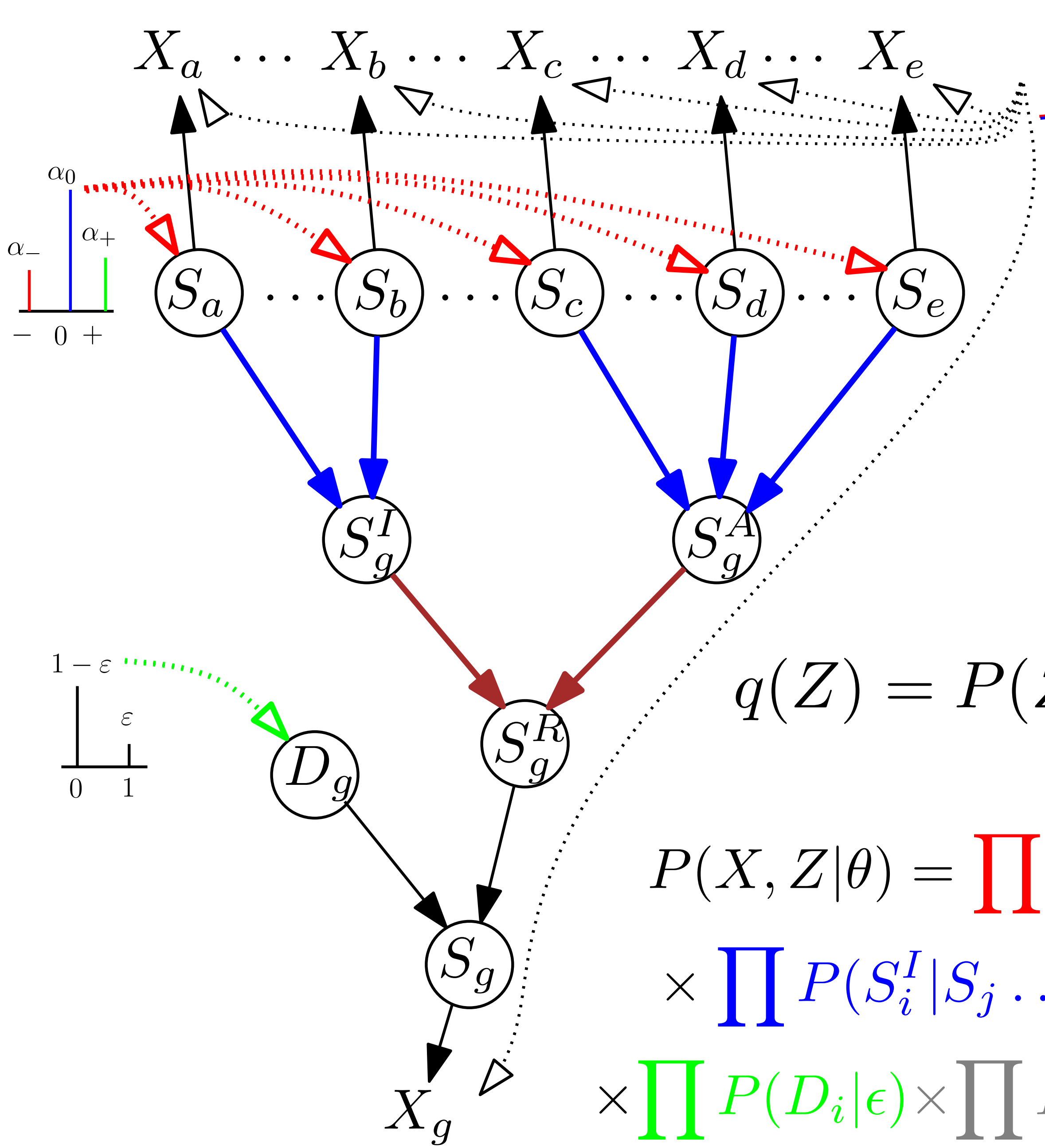


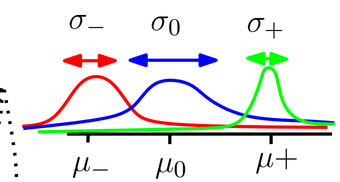
$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X,\theta) \sim P(Z,X|\theta)$$

$$P(X,Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots) \times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A) \times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

Detailed model : A Factor Graph



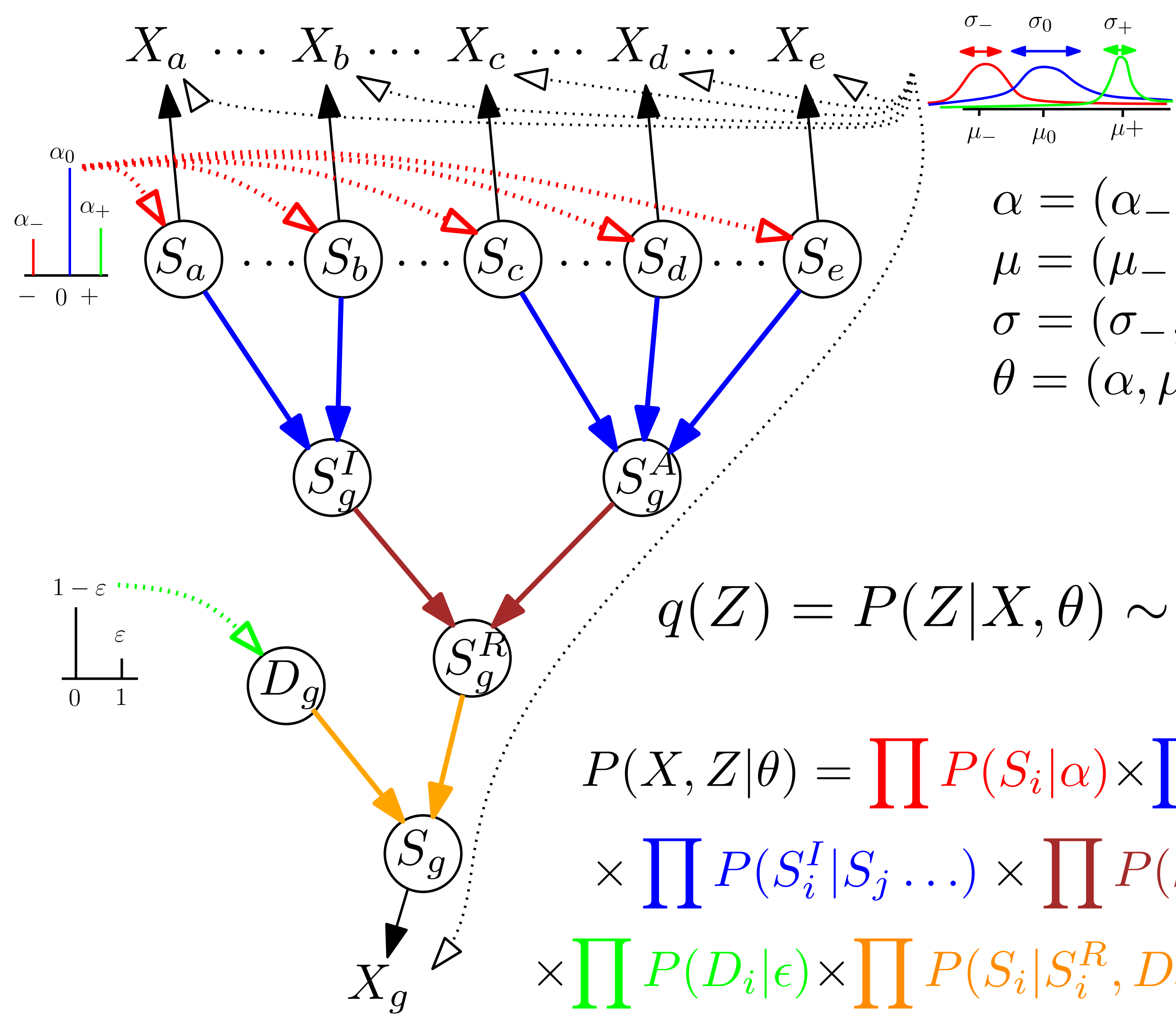


$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X,\theta) \sim P(Z,X|\theta)$$

$$P(X,Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots)$$
$$\times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A)$$
$$\times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

Detailed model : A Factor Graph

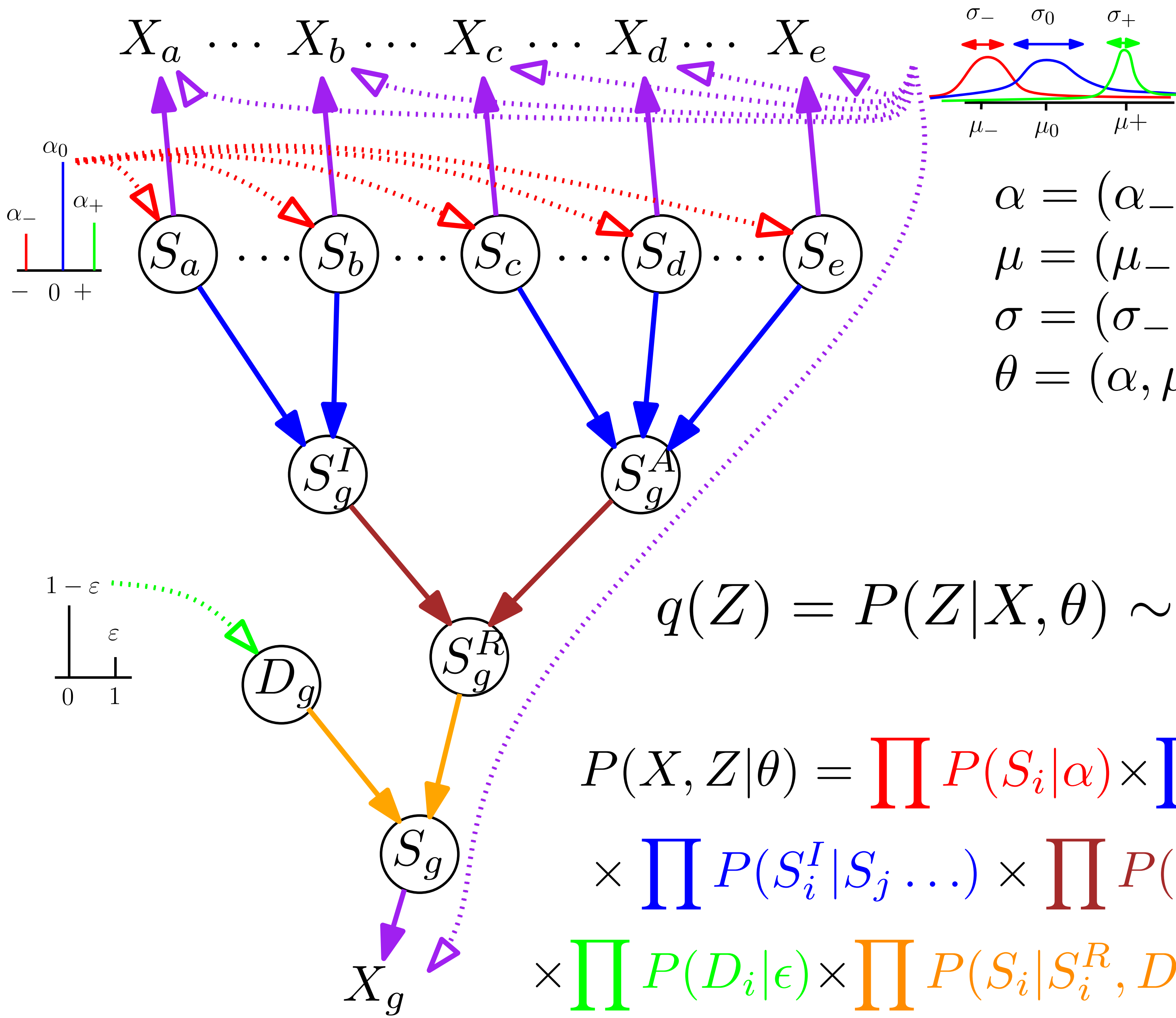


$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X,\theta) \sim P(Z,X|\theta)$$

$$P(X,Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots) \times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A) \times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

Detailed model : A Factor Graph

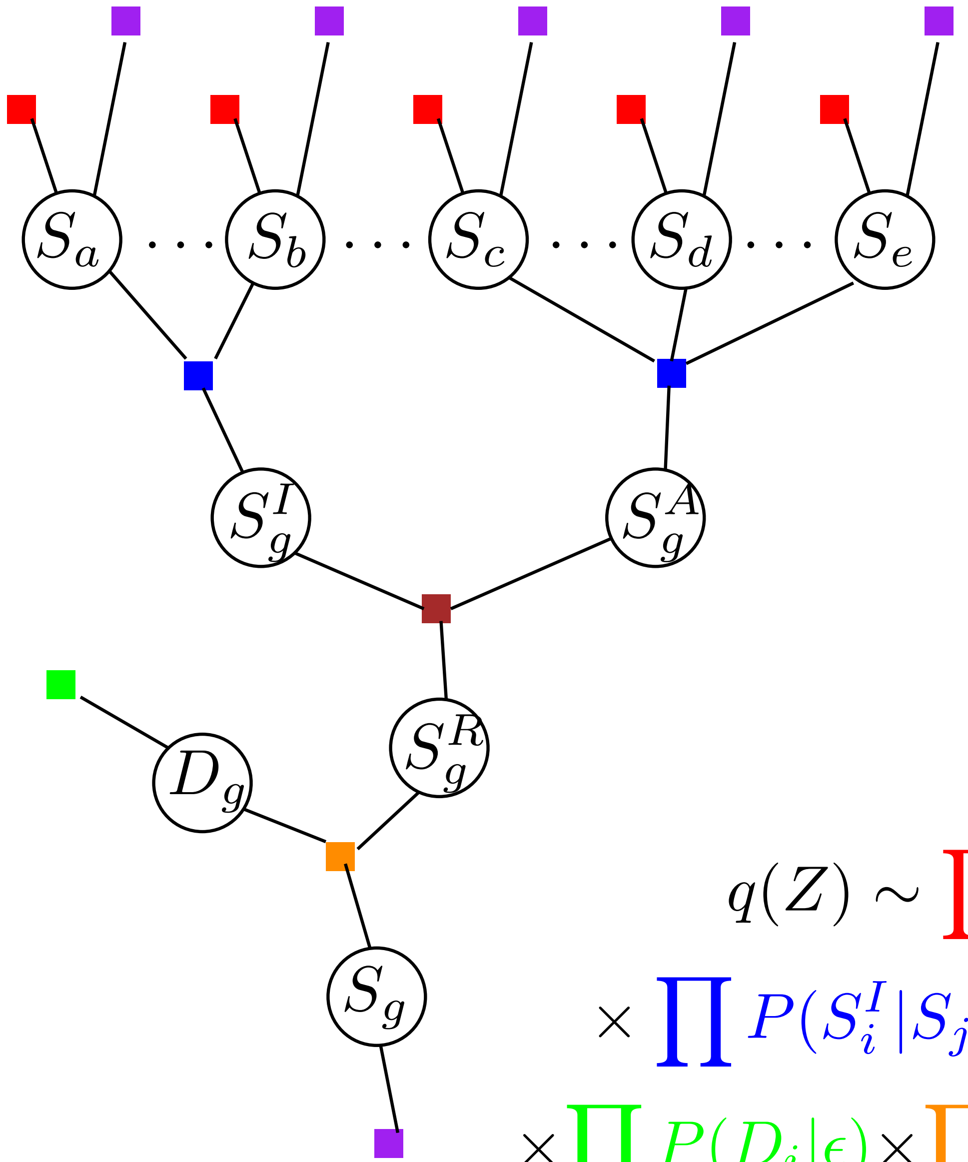


$$\alpha = (\alpha_-, \alpha_0, \alpha_+)$$
$$\mu = (\mu_-, \mu_0, \mu_+)$$
$$\sigma = (\sigma_-, \sigma_0, \sigma_+)$$
$$\theta = (\alpha, \mu, \sigma, \epsilon)$$

$$q(Z) = P(Z|X,\theta) \sim P(Z,X|\theta)$$

$$P(X,Z|\theta) = \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots) \times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A) \times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i,\mu,\sigma)$$

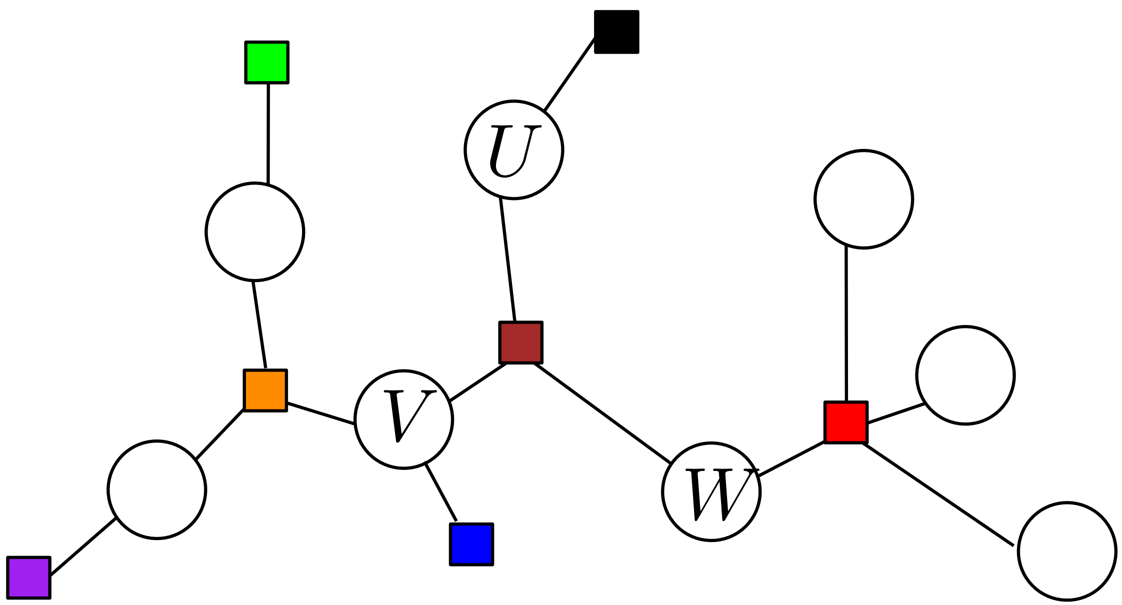
Detailed model : A Factor Graph



$$q(Z) \sim \prod P(S_i|\alpha) \times \prod P(S_i^A|S_j \ldots) \times \prod P(S_i^I|S_j \ldots) \times \prod P(S_i^R|S_i^I, S_i^A) \times \prod P(D_i|\epsilon) \times \prod P(S_i|S_i^R, D_i) \times \prod P(X_i|S_i, \mu, \sigma)$$

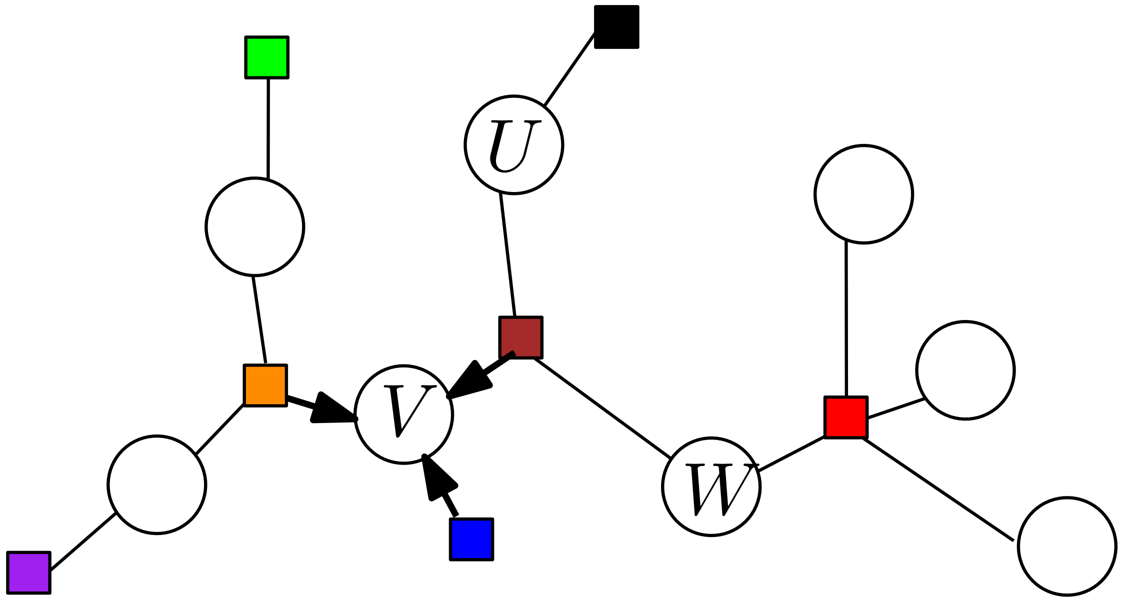
Solving Factor graphs: Belief Propagation

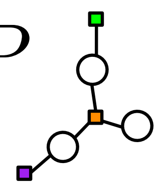
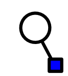
G=

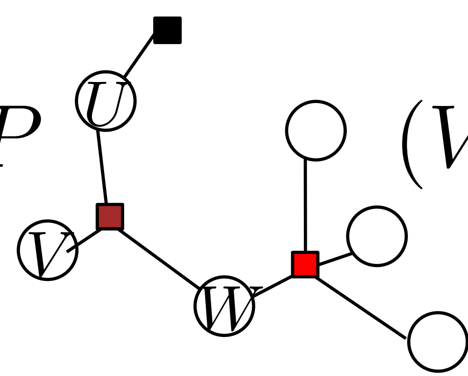


Solving Factor graphs: Belief Propagation

G=

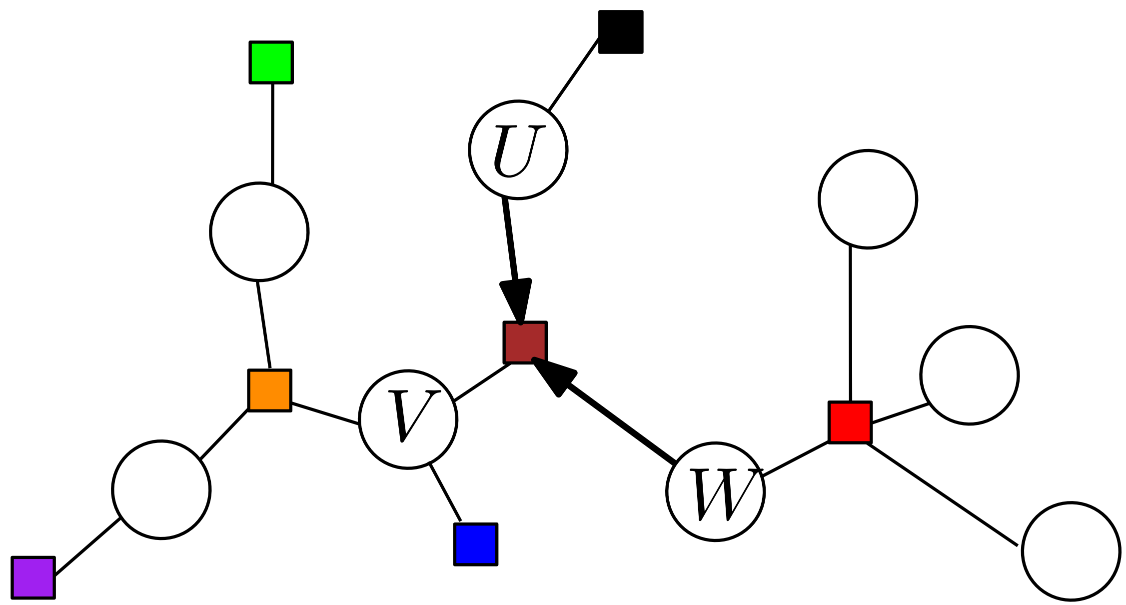


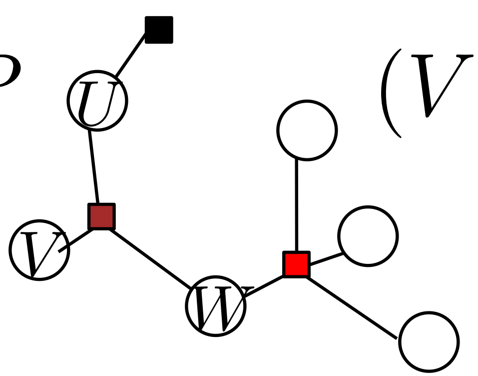
$P_G(V = v) = P$  $(V = v) \times P$  $(V = v)$

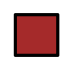
$\times P$ 

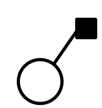
$(V = v)$

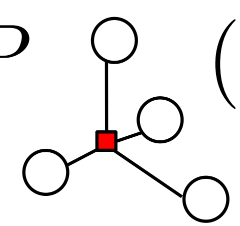
Solving Factor graphs: Belief Propagation

G= 

P 

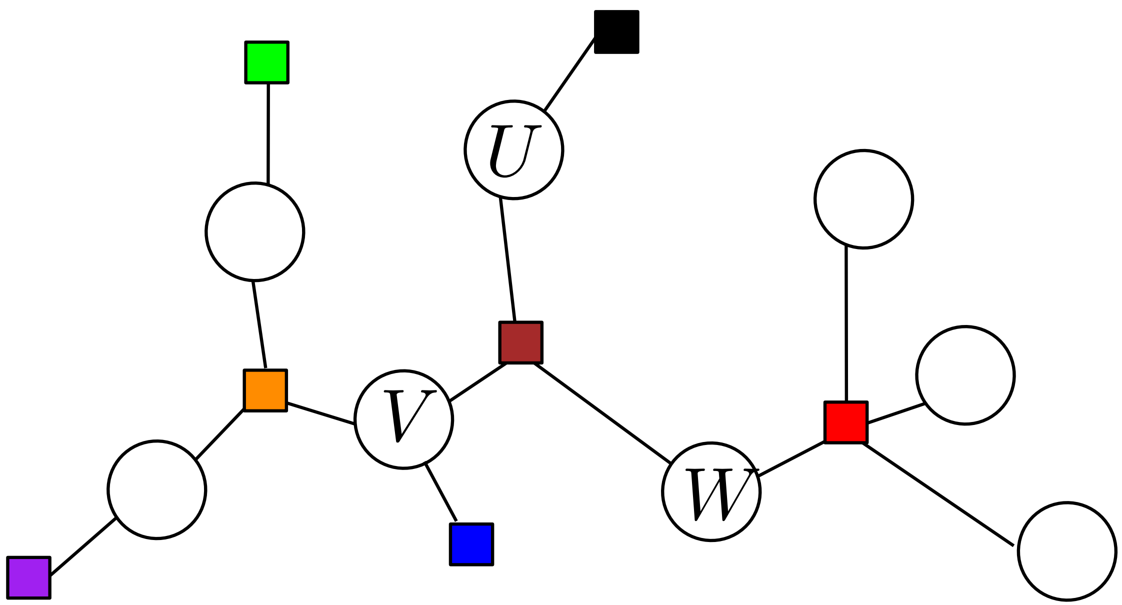
$(V = v) = \sum_{u,w}$  (u, v, w)

$\times P$  $(U = u)$

$\times P$  $(W = w)$

Solving Factor graphs: Belief Propagation

G=



- Belief Propagation computes the exact marginals in one linear pass when the factor graph is a tree
- If it has bounded tree-width, the exact marginals can be obtained working on a tree decomposition
- Otherwise, the propagation process must be iterated
- Theory shows it can go wrong, but experience shows it often converges towards the correct marginals.

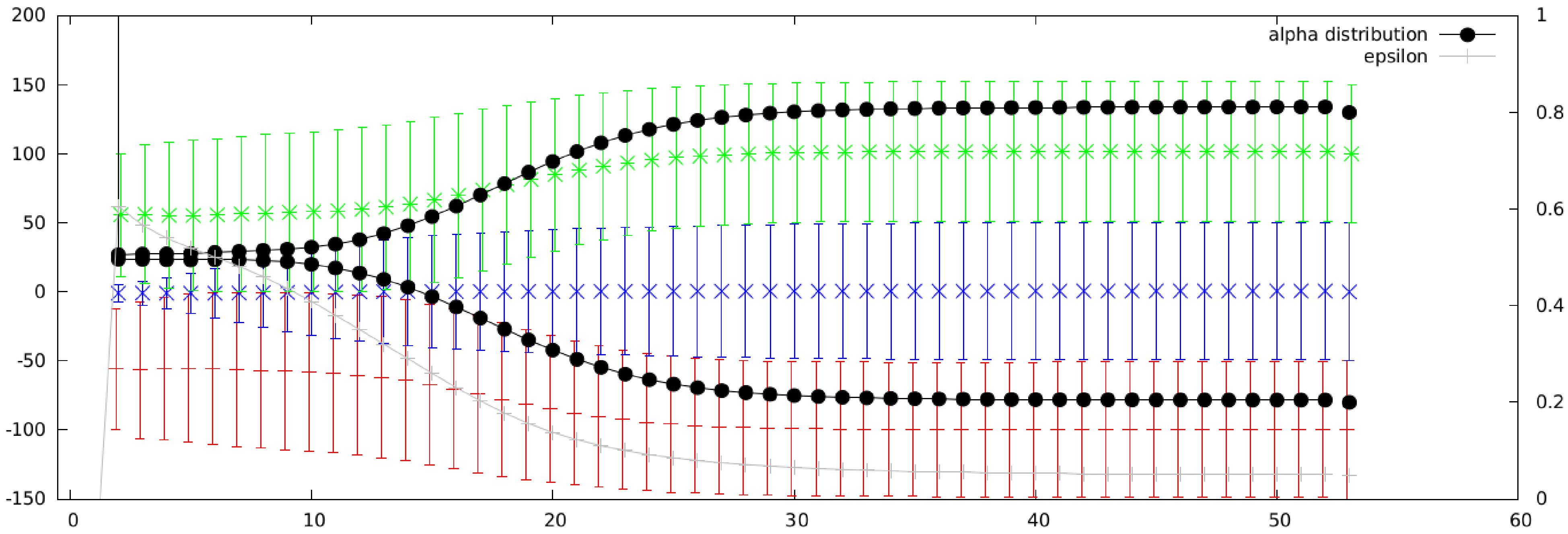
Summing up

- Using Belief Propagation, we obtain an approximation of q
- The EM algorithm uses this as an approximated step E
- The system parameters are inferred
- We obtain the posterior probability for each gene to be deregulated.

Validation ?

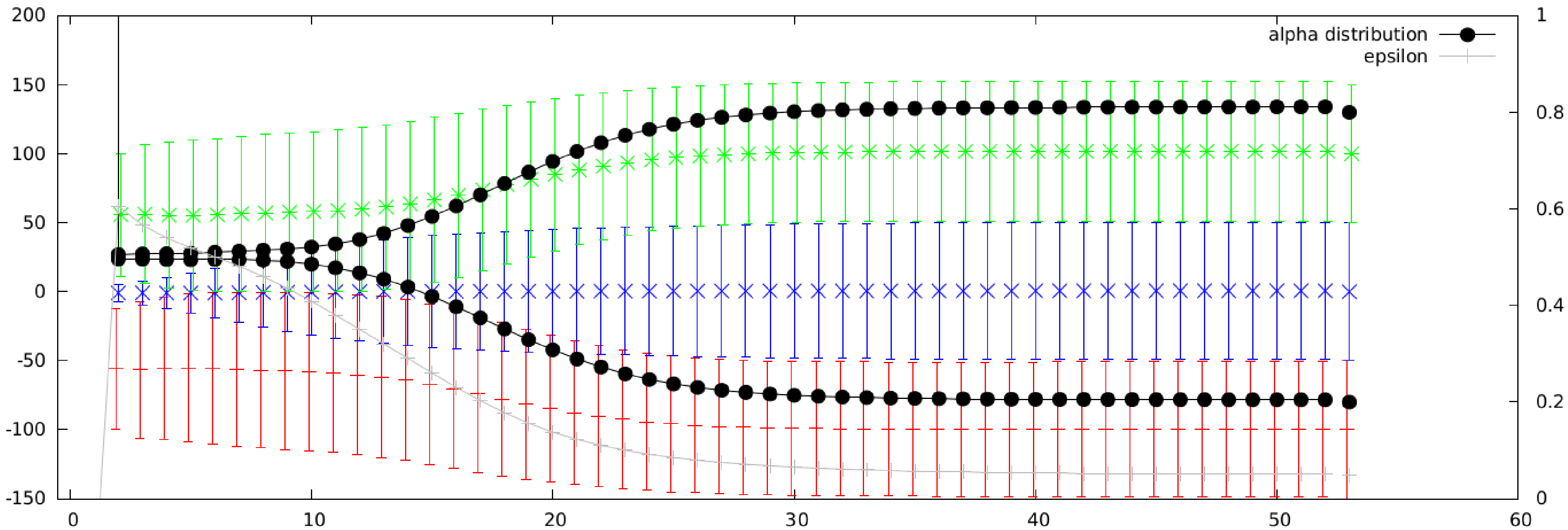
- No way to know the deregulation status in a real data set
- The method is tested on simulated data
- It is run on real data, and results are compared with other perturbations (Copy-Number Alterations)
- Further analyses of the results should give information about the tumors (cause, subtypes...)

EM progress



Small sensitivity to initial parameter set θ_0. (here $\mu_0 = (-1, 0, 1)$ and $\sigma_0 = (1, 1, 1)$)
First iteration usually finds good estimates of μ, σ, then α and ε are slowly refined.

EM progress



Small sensitivity to initial parameter set θ_0. (here $\mu_0 = (-1, 0, 1)$ and $\sigma_0 = (1, 1, 1)$)
First iteration usually finds good estimates of μ, σ, then α and ε are slowly refined.

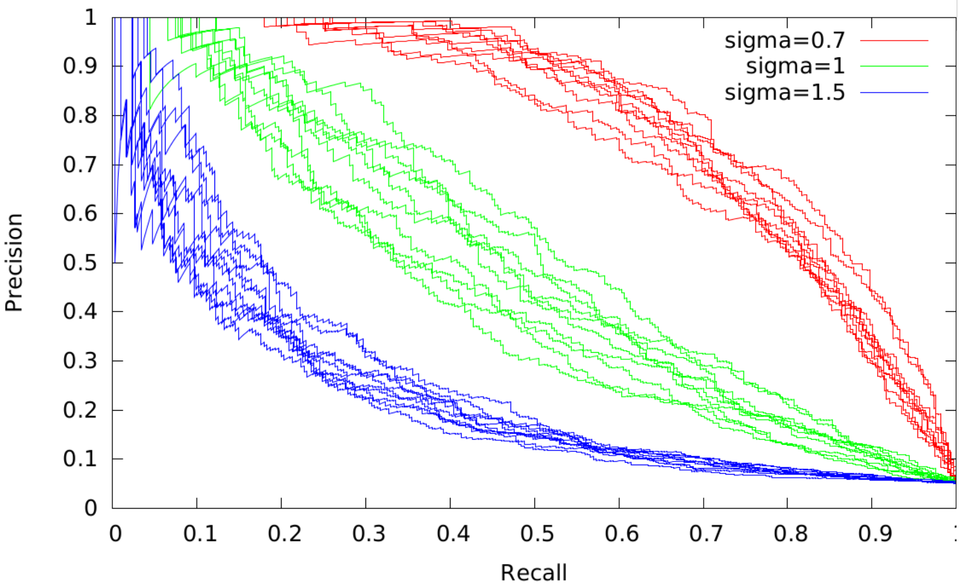
What about deregulated genes ?

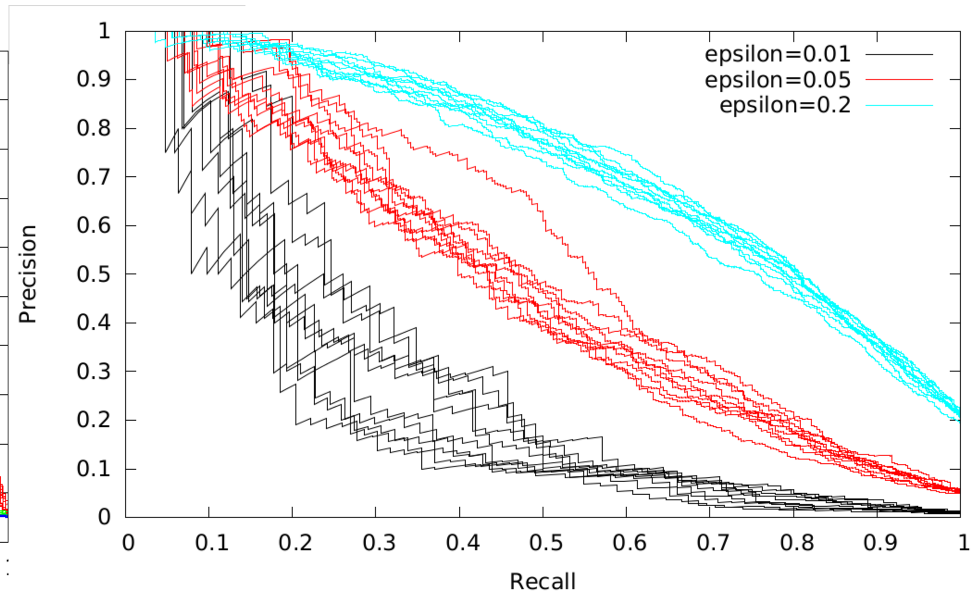
Precision-Recall (PR) curves

Default parameter set :

- $\mu = (-1, 0, 1)$
- $\sigma = (1, 1, 1)$
- $\alpha = (0.1, 0.8, 0.1)$
- $\epsilon = 0.1$

Then, various parameters are individually varied.



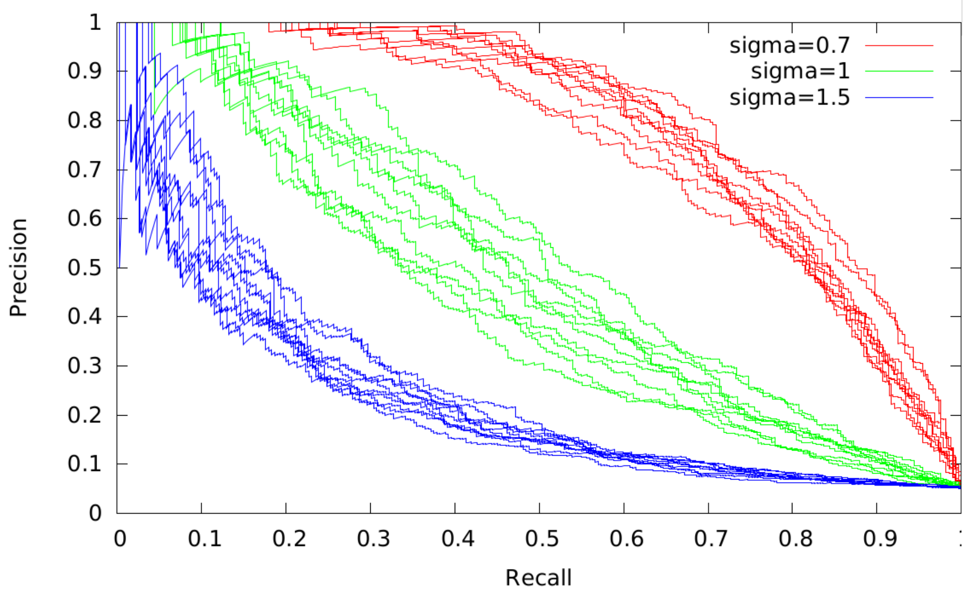


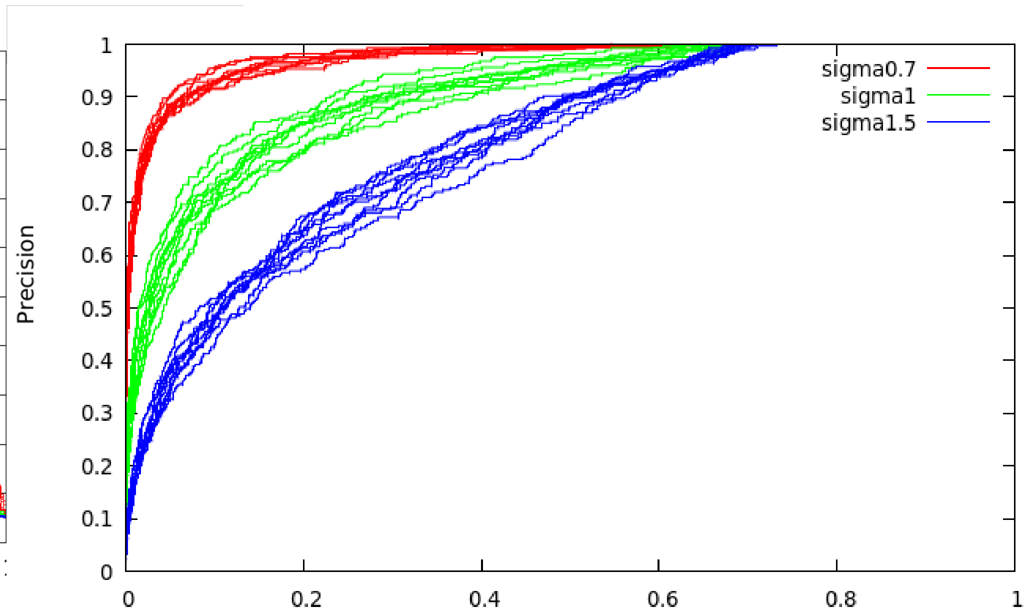
Precision-Recall (PR) curves

Default parameter set :

- $\mu = (-1, 0, 1)$
- $\sigma = (1, 1, 1)$
- $\alpha = (0.1, 0.8, 0.1)$
- $\epsilon = 0.1$

Then, various parameters are individually varied.



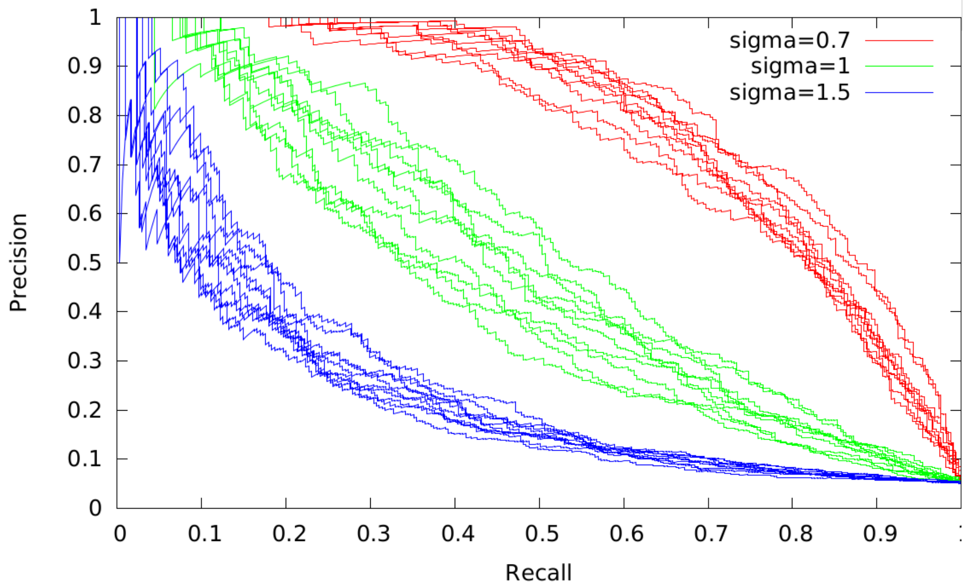


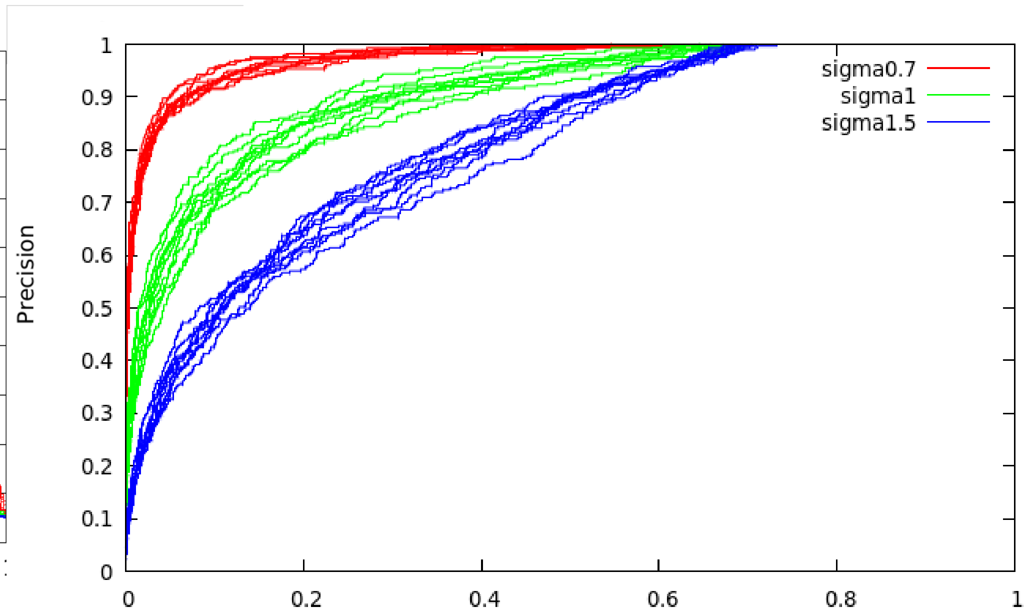
Precision-Recall (PR) curves

Default parameter set :

- $\mu = (-1, 0, 1)$
- $\sigma = (1, 1, 1)$
- $\alpha = (0.1, 0.8, 0.1)$
- $\epsilon = 0.1$

Then, various parameters are individually varied.





Thank you