Bayesian fragility modeling: study with

Straub, D. and Der Kiureghian, A. (2008), Improved seismic fragility modeling from empirical data, Structural Safety, 30, pp. 320-336.

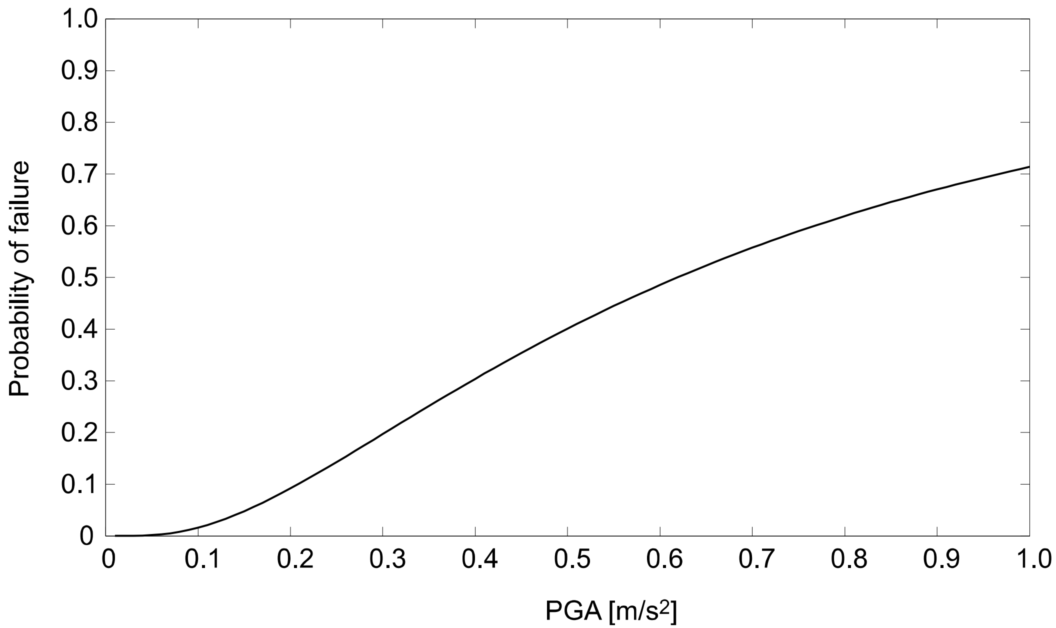
Kazuyoshi Nishijima
IBK, ETH Zurich

Contents

- A problem in classical fragility modeling using historical data
  → Problem addressed in Straub and Der Kiureghian (2008)
- Introduction to the approach proposed in Straub and Der Kiureghian (2008)
- Fragility modeling using Winbugs following the proposed approach

What is a fragility model?

Fragility model represents the **probability of failure** as a function of the value of **hazard index**.



The hazard index of earthquake:

e.g. peak ground acceleration (PGA) [m/s^2]

How to develop a fragility model?

- Empirical approaches
  → based on historical data and statistical analysis

- Engineering approaches
  → taking basis in engineering knowledge on structural behavior subject to loads.

Hybrid approaches have been proposed: e.g. first develop a fragility model by engineering approach; then update the model using historical data.

Problem in classical empirical approaches

Example of historical data

Table 1
Failure data for TR1 (1-phase 230 kV transformers)

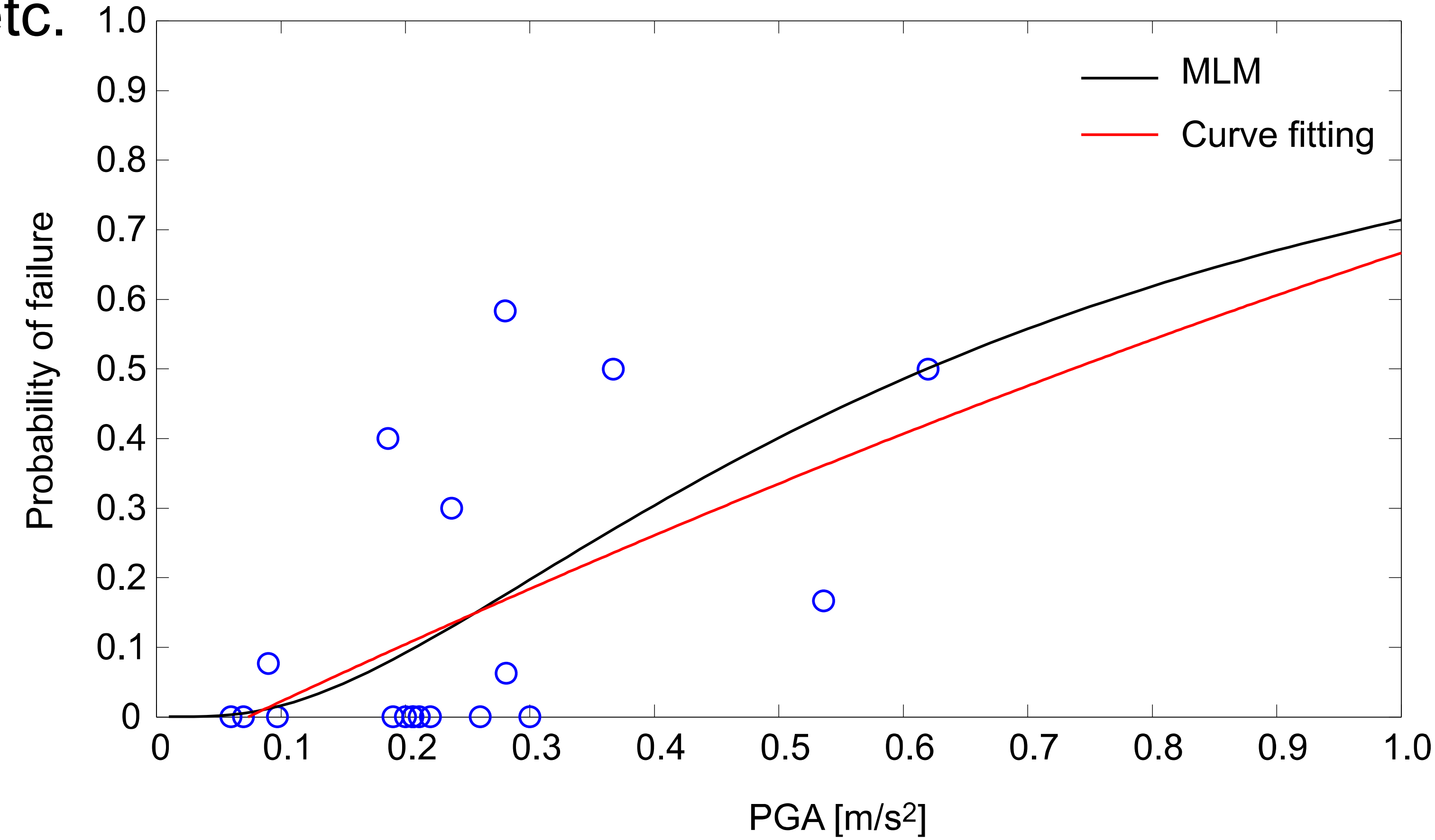
j	k	Date	$\hat{S}_{jk}$ (g)	n_{jk}	N_{jk}	n_{jk}/N_{jk}
1	1	1971	0.07	0	13	0.00
1	6	1987	0.09	1	13	0.08
2	6	1987	0.20	0	3	0.00
3	1	1971	0.211	0	7	0.00
3	6	1987	0.22	0	7	0.00
4	3	1983	0.30	0	3	0.00
5	2	1978	0.28	7	12	0.58
6	6	1987	0.19	0	6	0.00
7	1	1971	0.097	0	6	0.00
7	6	1987	0.26	0	6	0.00
8	4	1984	0.06	0	3	0.00
9	8	1989	0.206	0	9	0.00
10	8	1989	0.237	3	10	0.30
11	8	1989	0.281	1	16	0.06
12	8	1989	0.206	0	9	0.00
13	6	1987	0.62	5	10	0.50
14	9	1994	0.367	2	4	0.50
15	8	1989	0.186	4	10	0.40
16	1	1971	0.536	1	6	0.17

j: Substation, k: earthquake, $\hat{S}_{jk}$: estimated PGA, n_{jk}: number of failures, and N_{jk}: number of TR1 equipment in substation j during k.

Table 1. in Straub, D and Der Kiureghian, A (2008), Improved seismic fragility modeling from empirical data, Structural Safety, 30, pp. 320-336.

Classical empirical approaches

- Regression analyses/curve fitting
- Maximum likelihood methods (MLMs)
- etc.



Problem in classical empirical approaches

Usually, the values of hazard index are not directly observed.

Thus, the values utilized in the statistical analysis for fragility modeling are **estimated** values of hazard index;

→ i.e. they may not be true values, and thus associate uncertainties.

The classical empirical approaches may not consider such uncertainties appropriately.

Approach proposed in Straub and Der Kiureghian (2008)

Improved fragility modeling

Focusing on seismic fragility (yet the underlying idea is general), Straub and Der Kiureghian (2008)) propose

“an appropriate representation of the uncertainties in the estimation of fragility”.

→ This introduces statistical dependence among observations.

Statistical dependence through common factors

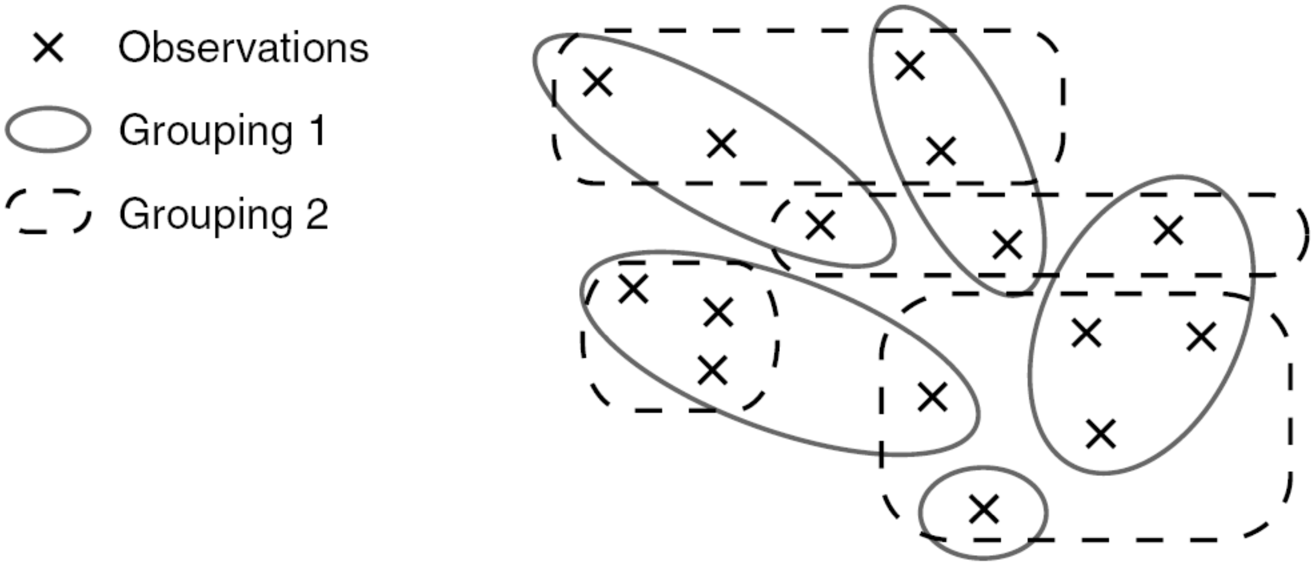


Fig. 1. Illustration of grouping of observations (from Straub and Der Kiureghian (2008))

Fragility model for electrical substation equipment

The assumptions:

(a) Observations for equipment are exchangeable.

(b) Estimated PGAs are subject to estimation errors.

(c) PGA is selected as the scalar hazard index.

(d) Component capacities are: independent in different substations; dependent within substations due to common effects, e.g. age.

(e) Modeling errors in describing seismic demand are: independent in different substations in different earthquake; dependent within substations during the same earthquake

(f) No deterioration; replaced by new equipment upon failure.

Limit state function

The failure of the equipment can be defined through:

$$g_{ijk} = r_{ij} - x_{jk} \begin{cases} \geq 0 & \text{(no failure)} \\ < 0 & \text{(failure)} \end{cases}$$

where r_{ij} is the capacity for equipment i at substation j, and x_{jk} is the "effective" seismic demand at substation j during earthquake k. The effective seismic demand is defined as:

$$x_{jk} = \hat{s}_{jk} + \varepsilon_{jk} - y_{jk}$$

where $\hat{s}_{jk}$ is the estimated PGA, ε_{jk} is the estimation error, and y_{jk} represents the common effects in (d).

Assumed distributions

r_{ij} : is independent and identically distributed (iid) and follows the normal distribution with the mean μ_r and standard deviation σ_r ($N(\mu_r, \sigma_r{}^2)$).

ε_{jk} : is iid and follows $N(\mu_\varepsilon, \sigma_\varepsilon{}^2)$, where $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 0.3$.

y_{jk} : is iid and follows $N(\mu_y, \sigma_y{}^2)$, where $\mu_y = 0$.

The goal is to estimate the posterior distributions of μ_r ,σ_r and σ_y .

Likelihood function

Unknown parameters: $\boldsymbol{\theta} = (\mu_r, \sigma_r, \sigma_y)$

Data: $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_j, ..., \mathbf{z}_M)$

Likelihood function: $L(\boldsymbol{\theta} \mid \mathbf{z})$

The likelihood function can be written as:

$$L(\boldsymbol{\theta} \mid \mathbf{z}) = \prod_{j=1}^{M} L(\boldsymbol{\theta} \mid \mathbf{z}_j)$$

and the components of the likelihood function can be written as:

$$L(\boldsymbol{\theta} \mid \mathbf{z}_j) = \int \prod_i P(z_{ij} \mid \mathbf{x}_j) f(\mathbf{x}_j) d\mathbf{x}_j$$

where $\mathbf{x}_j$ and $\mathbf{z}_j$ are the collections of x_{jk} and z_{ij} respectively.

Likelihood function

(1) For substation subject to only single earthquake k,

$$L(\boldsymbol{\theta} \mid \mathbf{z}_j) = \int \prod_i P(z_{ijk} \mid x_{jk}) f(x_{jk}) dx_{jk}$$

(1-1) For components that failed:

$$P(z_{ijk} \mid x_{jk}) = \Phi\left(\frac{x_{jk} - \mu_r}{\sigma_r} \right)$$

(1-2) For components that survived:

$$P(z_{ijk} \mid x_{jk}) = \Phi\left(-\frac{x_{jk} - \mu_r}{\sigma_r} \right)$$

Likelihood function

(2) For substation subject to two earthquakes k (first) and l (second):

$$L(\boldsymbol{\theta} \mid \mathbf{z}_j) = \int \prod_i P(z_{ijk} \mid x_{jk}, x_{jl}) f(x_{jk}) f(x_{jl}) dx_{jk} dx_{jl}$$

(2-1) For components that survived both earthquakes:

$$P(z_{ijk} \mid x_{jk}) = \Phi\left(-\frac{\max(x_{jk}, x_{jl}) - \mu_r}{\sigma_r} \right)$$

Likelihood function

(2-2) For components that survived the first earthquake and failed the second earthquake:

$$P(z_{ijk} \mid x_{jk}) = \begin{cases} \Phi\left(\dfrac{x_{jl} - \mu_r}{\sigma_r}\right) - \Phi\left(\dfrac{x_{jk} - \mu_r}{\sigma_r}\right) & (x_{jl} > x_{jk}) \\ 0 & (x_{jl} \le x_{jk}) \end{cases}$$

(2-3) For components that failed the first earthquake and survived the second earthquake:

$$P(z_{ijk} \mid x_{jk}) = \Phi\left(\frac{x_{jk} - \mu_r}{\sigma_r}\right) \cdot \Phi\left(-\frac{x_{jl} - \mu_r}{\sigma_r}\right)$$

Likelihood function

(2-4) For components that failed both earthquakes:

$$P(z_{ijk} \mid x_{jk}) = \Phi\left(\frac{x_{jk} - \mu_r}{\sigma_r}\right) \cdot \Phi\left(\frac{x_{jl} - \mu_r}{\sigma_r}\right)$$

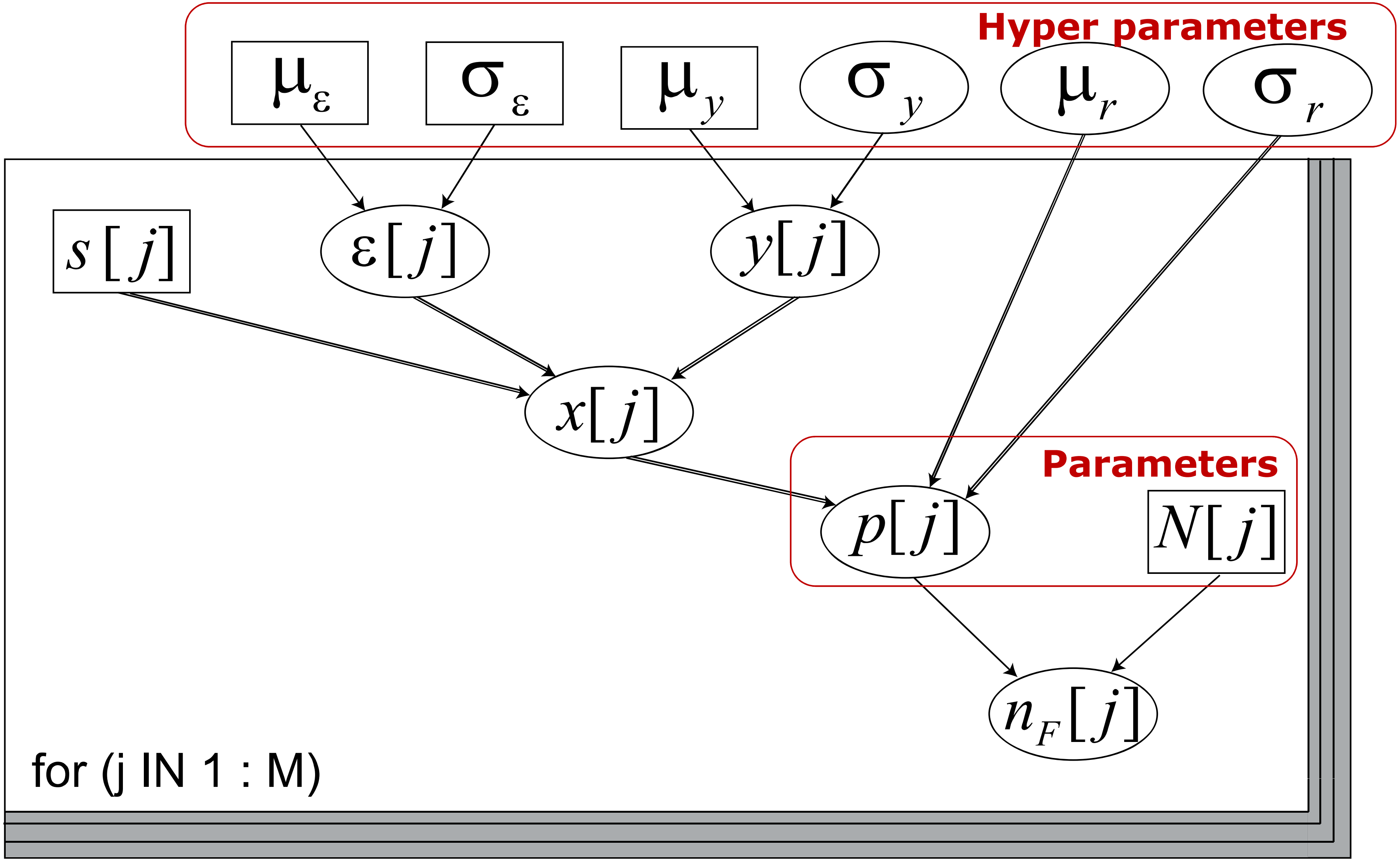
Fragility modeling using Winbugs

Case: substations subject to single earthquake

Start with a simplistic case: i.e.

data with substations subject to single earthquake.

Graphical representation



Full joint probability

$$
\begin{aligned}
&P[\mathbf{n}_F,\mathbf{p},\mathbf{x},\boldsymbol{\varepsilon},\mathbf{y},\sigma_y,\mu_r,\sigma_r] \\
&= P[\mathbf{n}_F \mid \mathbf{p}]P[\mathbf{p} \mid \mathbf{x},\mu_r,\sigma_r]P[\mathbf{x} \mid \boldsymbol{\varepsilon},\mathbf{y}]P[\mathbf{y} \mid \sigma_y]P[\boldsymbol{\varepsilon}]P[\sigma_y]P[\mu_r]P[\sigma_r] \\
&= P[\sigma_y]P[\mu_r]P[\sigma_r]\cdot\left[\prod_j \binom{N_j}{n_{Fj}} p_j^{\,n_{Fj}}(1-p_j)^{N_j-n_{Fj}}\,\delta\left(p_j-\Phi\left(\frac{x_j-\mu_r}{\sigma_r}\right)\right)\right. \\
&\quad \left.\times\delta\left(x_j-s_j-\varepsilon_j+y_j\right)\frac{1}{\sigma_y}\phi\left(\frac{y_j-\mu_y}{\sigma_y}\right)\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_j-\mu_\varepsilon}{\sigma_\varepsilon}\right)\right]
\end{aligned}
$$

where

$\delta(\cdot)$: delta function

$\phi(\cdot)$: standard normal probability density function.

Posterior probability

$$P[\sigma_y, \mu_r, \sigma_r \mid \mathbf{n}_F]$$

$$= \int P[\sigma_y]P[\mu_r]P[\sigma_r] \cdot \left[\prod_j \binom{N_j}{n_{Fj}} p_j^{n_{Fj}} (1-p_j)^{N_j - n_{Fj}} \delta\left(p_j - \Phi\left(\frac{x_j - \mu_r}{\sigma_r} \right) \right) \right.$$

$$\left. \times \delta\left(x_j - s_j - \varepsilon_j + y_j \right) \frac{1}{\sigma_y} \phi\left(\frac{y_j - \mu_y}{\sigma_y} \right) \frac{1}{\sigma_\varepsilon} \phi\left(\frac{y_j - \mu_\varepsilon}{\sigma_\varepsilon} \right) \right] d\mathbf{p} d\mathbf{x} d\boldsymbol{\varepsilon} d\mathbf{y}$$

$$= P[\sigma_y]P[\mu_r]P[\sigma_r] \int \cdot \left[\prod_j \binom{N_j}{n_{Fj}} \Phi\left(\frac{s_j + \varepsilon_j - y_j - \mu_r}{\sigma_r} \right)^{n_{Fj}} \right.$$

$$\left. \times \left(1 - \Phi\left(\frac{s_j + \varepsilon_j - y_j - \mu_r}{\sigma_r} \right) \right)^{N_j - n_{Fj}} \frac{1}{\sigma_y} \phi\left(\frac{y_j - \mu_y}{\sigma_y} \right) \frac{1}{\sigma_\varepsilon} \phi\left(\frac{y_j - \mu_\varepsilon}{\sigma_\varepsilon} \right) \right] d\boldsymbol{\varepsilon} d\mathbf{y}$$

Posterior probability

$$P[\sigma_y, \mu_r, \sigma_r \mid \mathbf{n}_F]$$

$$= P[\sigma_y] P[\mu_r] P[\sigma_r]$$

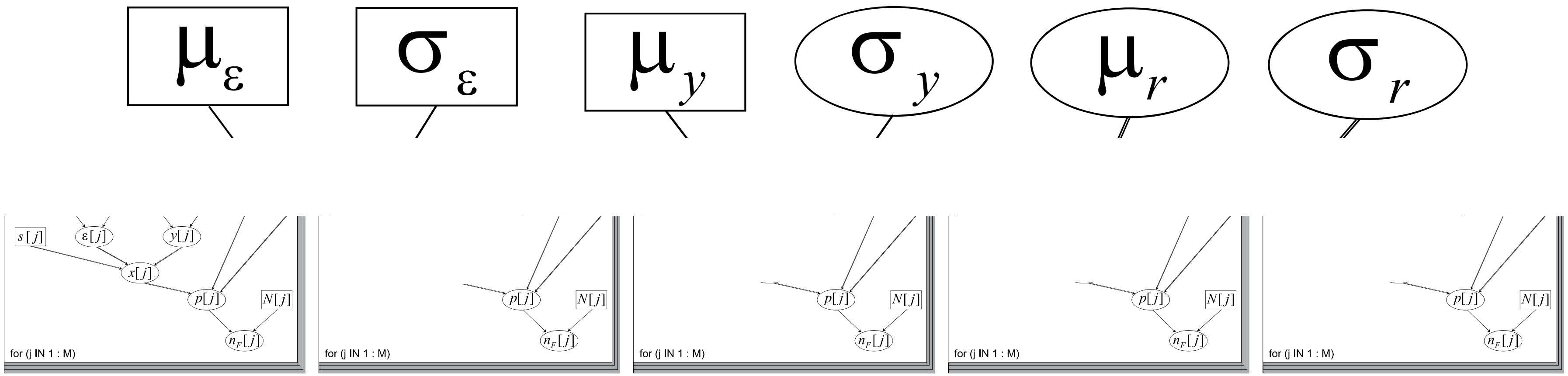
$$\times \prod_j \int \Bigg[\binom{N_j}{n_{Fj}} \Phi\left(\frac{s_j + \varepsilon_j - y_j - \mu_r}{\sigma_r}\right)^{n_{Fj}} \left(1 - \Phi\left(\frac{s_j + \varepsilon_j - y_j - \mu_r}{\sigma_r}\right)\right)^{N_j - n_{Fj}}$$

$$\times \frac{1}{\sigma_y} \phi\left(\frac{y_j - \mu_y}{\sigma_y}\right) \frac{1}{\sigma_\varepsilon} \phi\left(\frac{y_j - \mu_\varepsilon}{\sigma_\varepsilon}\right) \Bigg] d\varepsilon_j dy_j$$

This is equivalent to Equations (8-12) in Straub and Der Kiureghian (2008) in the case of single earthquake.

Extension to the case of multiple earthquakes

Idea: developing networks for individual cases separately, however, sharing the hyper parameters (see the example Winbugs code)



Case (1) Case (2-1) Case (2-2) Case (2-3) Case (2-4)