Using Hybrid Dynamic Bayesian Networks to model Operational Risk in Finance

Prof Martin Neil
Queen Mary, University of London
& Agena Ltd
London, UK

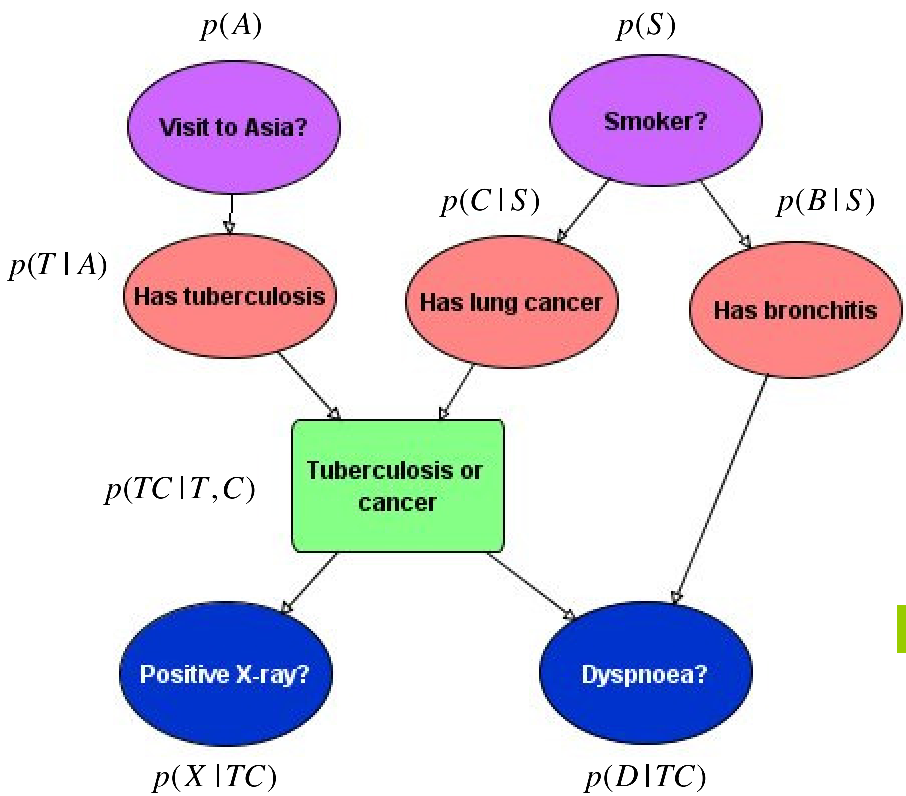
Contents

- Overview of Bayesian Networks
  - Discrete/Hybrid BNs
  - Dynamic discretization
- Operational Risk Application
  - Resiliency perspective
  - Rogue Trading

Bayesian networks

Use DAG to factorise Joint Distribution

Full probability model specified by graph and by conditional probabilities:

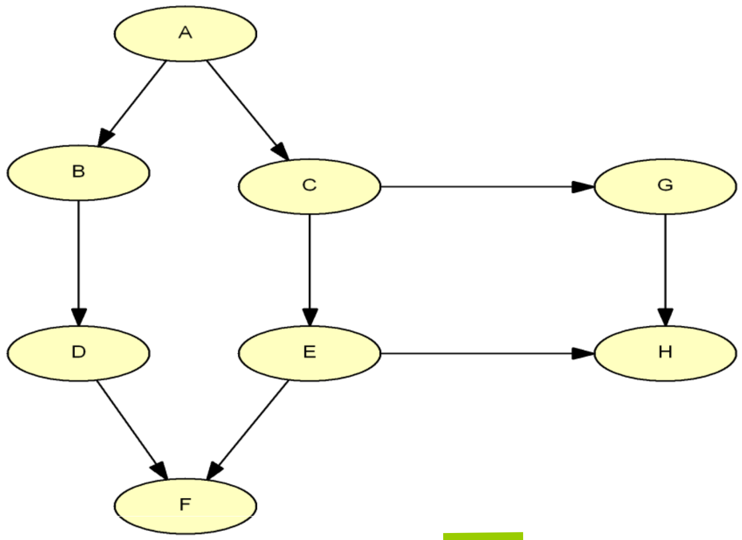


$$p(X_1, \ldots, X_d) = \prod_{i=1}^{d} p\left(X_i \middle| pa(X_i)\right)$$

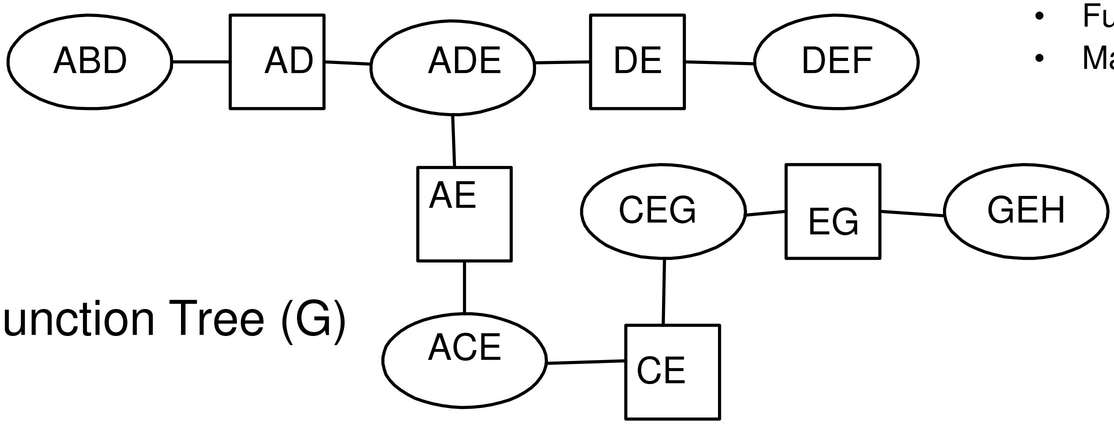
$$p(A, S, T, C, B, TC, X, D) =$$
$$p(A)\,p(T \mid A)\,p(S)\,p(C \mid S)\,p(B \mid S)$$
$$p(TC \mid T, C)\,p(X \mid TC)\,p(D \mid TC)$$

From $2^8 - 1 = 255$ parameters to $1+2+1+2+2+4+2+4 = 18$

Step 1: Identify Cliques and Create Junction Tree



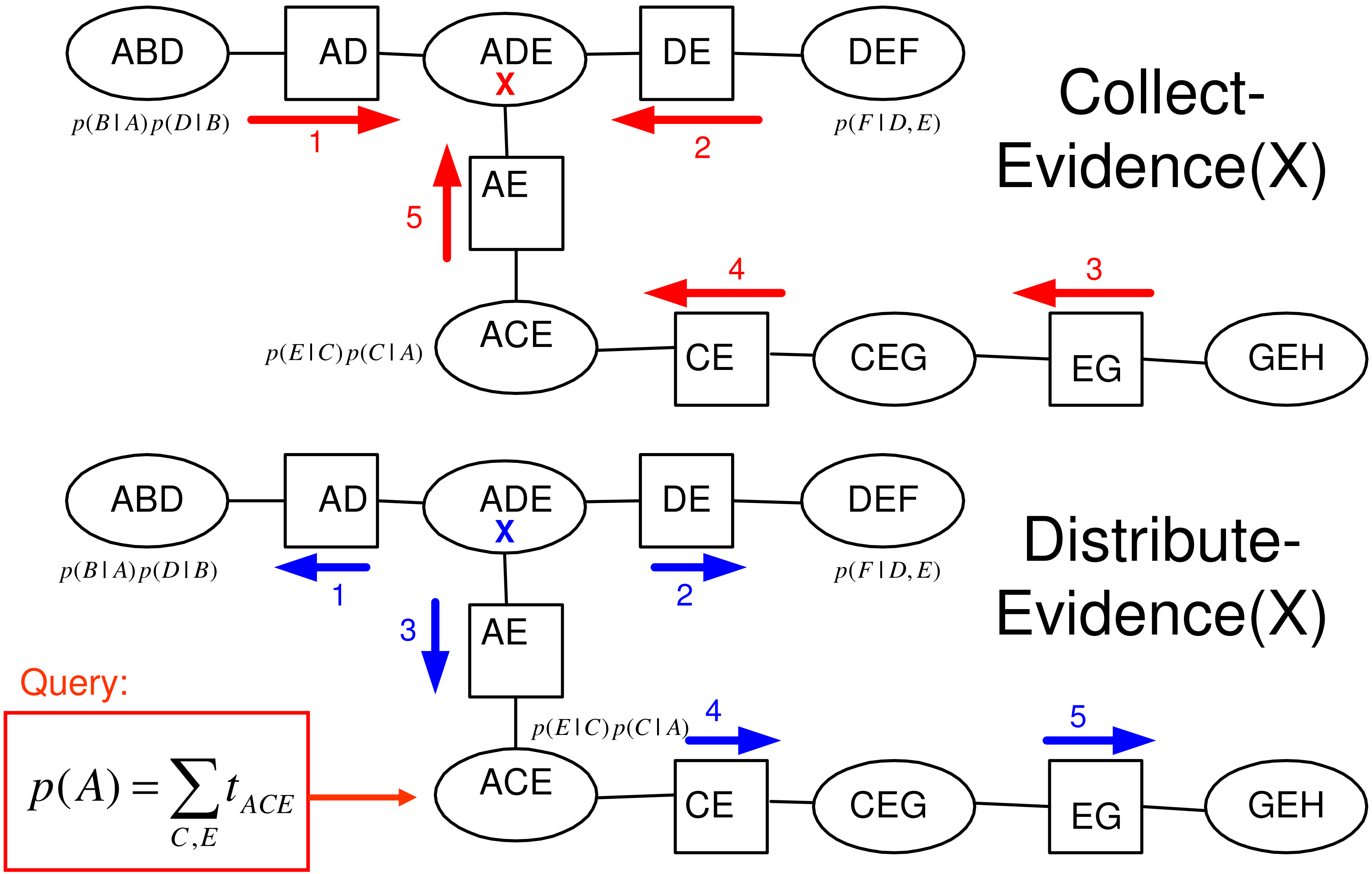
BN (G)



Junction Tree (G)

- Aim to compute posterior marginal distributions for variables of interest conditioned on evidence/data
- Methods:
  - Variable elimination or message passing algorithms: JunctionTree, PolyTree, Shenoy-Shafer etc.
- Propagate evidence through junction tree
  - When evidence entered calculate changes to BN locally
  - Propagate impact of evidence globally through junction tree
  - Use message passing to update likelihoods throughout BN
  - Calculations done using:
    - Bayes theorem (table multiplication and modular calculation)
    - Fundamental rule (table division)
    - Marginalisation (table factoring)

Step Two: Propagation – Collect and Distribute

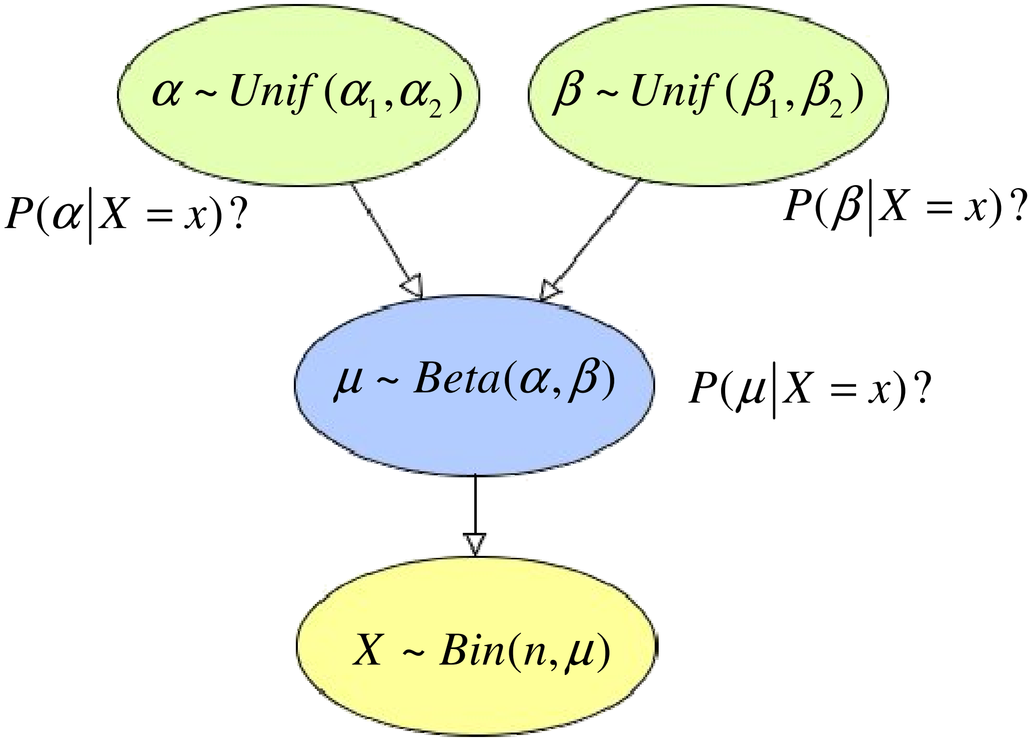


Properties and Problems

- Inference in discrete models is:
  - Exact
  - Modular
  - NP hard in principle but in practice many models are feasible with acceptable performance
- If all variables are continuous then use analytical solutions (assuming conjugacy) or approximations
- But what about continuous variables in BNs?
    - Continuous Gaussian (CG) distributions
    - Uniform (static) discretisation

Properties and Problems

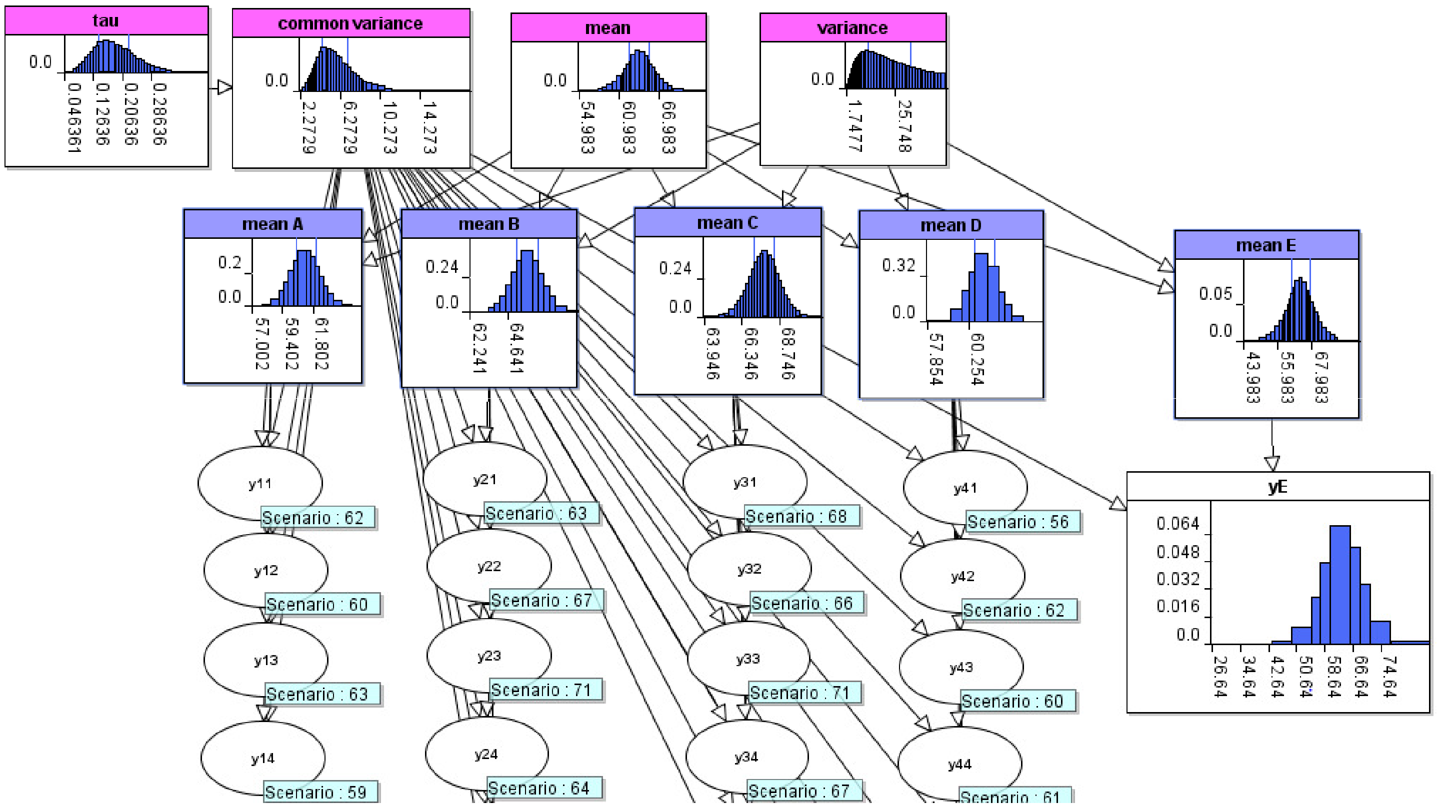
What about Bayesian statistical inference?



- (Near) Impossible in previous generation of BN algorithms
- E.g. for hybrid models

Need to use stochastic approximations: Monte Carlo Markov Chain (MCMC) – Gibbs sampling

Bayesian Hierarchical Models



- Two level hierarchical normal model as HBN
- Take four groups, A-D, learn hyper parameters from data
- Predict parameters for unknown, new group, E

Static Discretization

X Continuous valued node

$\Omega_X = [x_1, x_2],]x_1, x_2], \ldots,]x_{n-1}, x_n]$

Ω_S Cartesian product of $X \in S$ $\Omega_S = \otimes\{\Omega_X \mid X \in S\}$

$$P(X) = \sum_{\mathbf{y} \in \Omega_{S\text{-}X}} P(X, \mathbf{y})$$

Intervals of Ω_S defined in advance

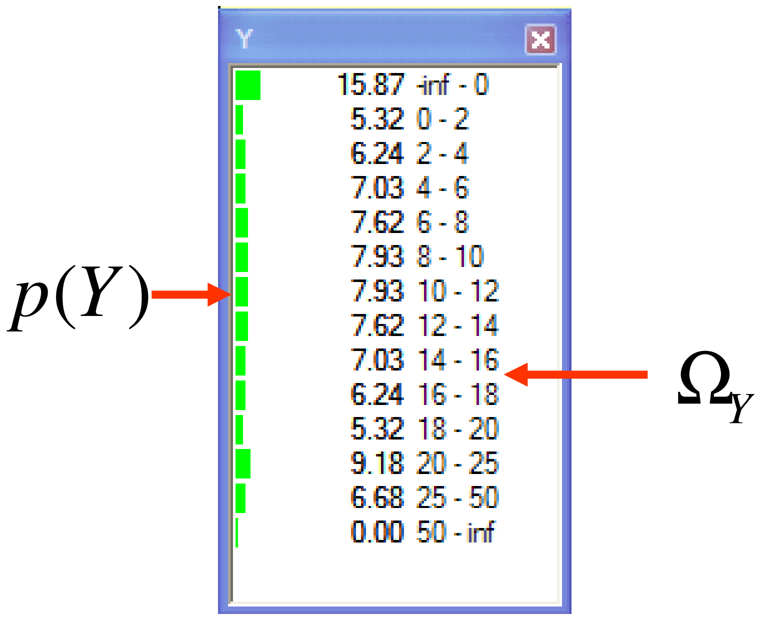
1. Discretize uniformly the domain Ω_X of the density function *f* as $\Omega_X = \bigcup \omega_i$
2. Define the discretized density function f_D as piecewise uniform (with constant value given by the mean value of *f* in each subregion)

Problems with Static Discretization

- The discretization is defined in advance and remains fixed throughout all subsequent stages regardless of any new conditional evidence
- High density regions (HDRs) need to be known in advance
- Trade-off between precision and computational cost

Example:

$$f(Y) = N(10,100)$$



Hugin

Solution - Dynamic Discretisation

Stage 1: Given a discretisation $\Omega = \bigcup \omega_i$ of the domain Ω of the density function *f* , define the discretised density function f_D as piecewise constant in each subregion ω_i (with constant given by the mean value of *f*)

Stage 2: Find a discretisation $\{\omega_i\}$ that minimises the relative entropy error

$$E_j = \left[\frac{f_{\max} - \bar{f}}{f_{\max} - f_{\min}} f_{\min} \log \frac{f_{\min}}{\bar{f}} + \frac{\bar{f} - f_{\min}}{f_{\max} - f_{\min}} f_{\max} \log \frac{f_{\max}}{\bar{f}} \right] |\omega_j|$$

* Upper bound approximation to Kullback-Leibler (KL) distance

Outline of DD Algorithm

- Convert BN to Junction Tree
- Choose initial discretisation $\Omega^{(0)}$ for all continuous variables
- FOR iterations $k = 1, 2, ..., n$
  - FOR all continuous nodes X_i

    Calculate $p^{(k)}(X_i \mid parents\{X\}, e)$ on $\Omega^{(k-1)}$

    Propagate model using collect() and distribute()

    Query model to get posterior marginal $p^{(k)}(X_i \mid e)$

    Calculate Sum Entropy Error E_k for $p^{(k)}(X_i \mid e)$

    IF E_k is acceptable STOP

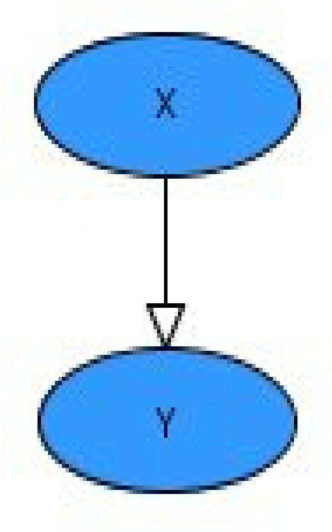
    ELSE create new $\Omega^{(k)}$ by splitting interval with $\max(E_j)$

  ENDFOR

ENDFOR

Example: Mixture of Gaussians

Conditional distributions:



$$p(X = false) = 0.5,\ \ p(X = true) = 0.5$$

$$p(Y \mid X) = \begin{cases} Normal(10,100) & X = false \\ Normal(50,10) & X = true \end{cases}$$

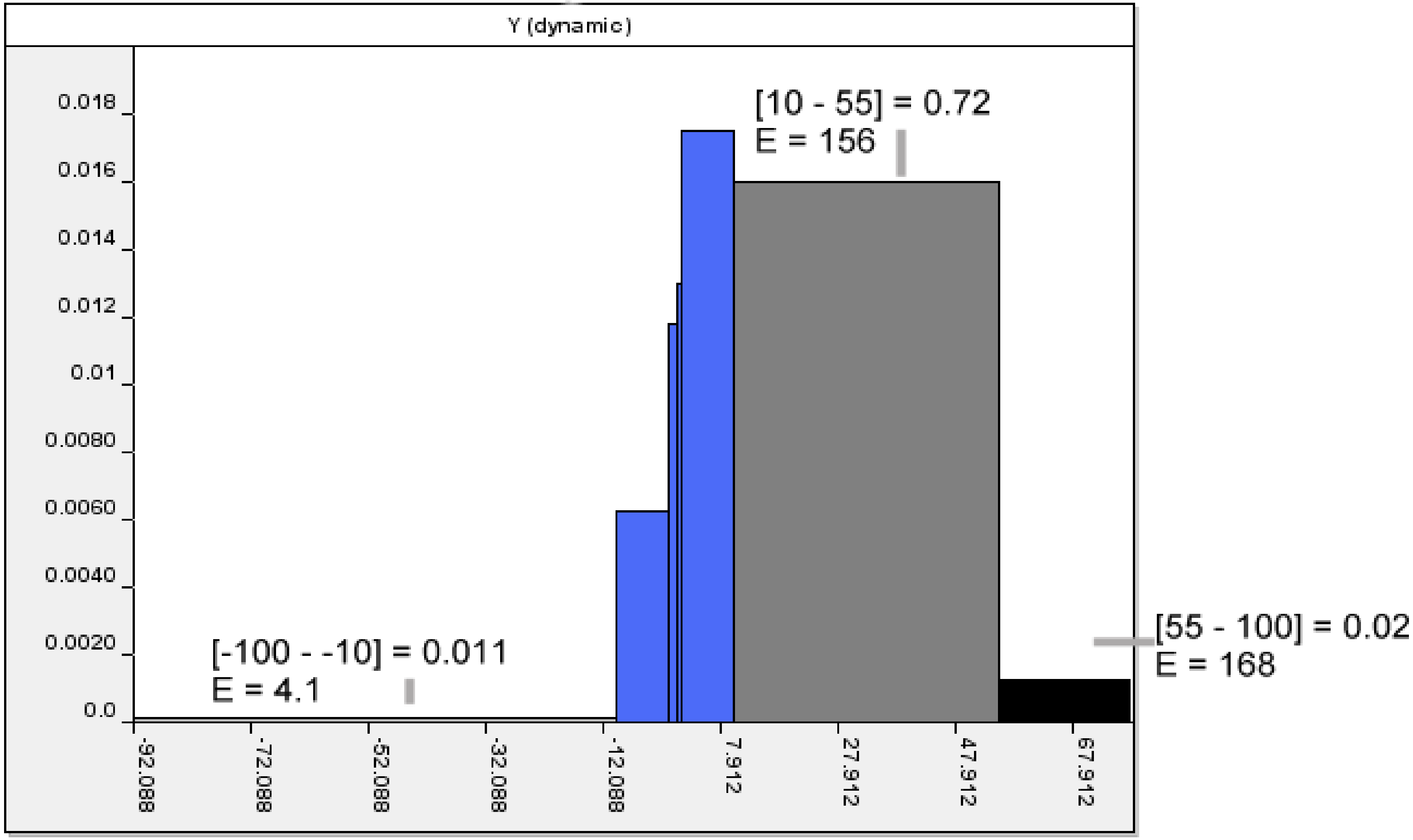
Marginal distribution:

$$p(Y) = \sum_X p\left(X\right) p(Y \middle| X) = \frac{1}{2}\left[Normal(y;10,100) + Normal(y;50,10)\right]$$

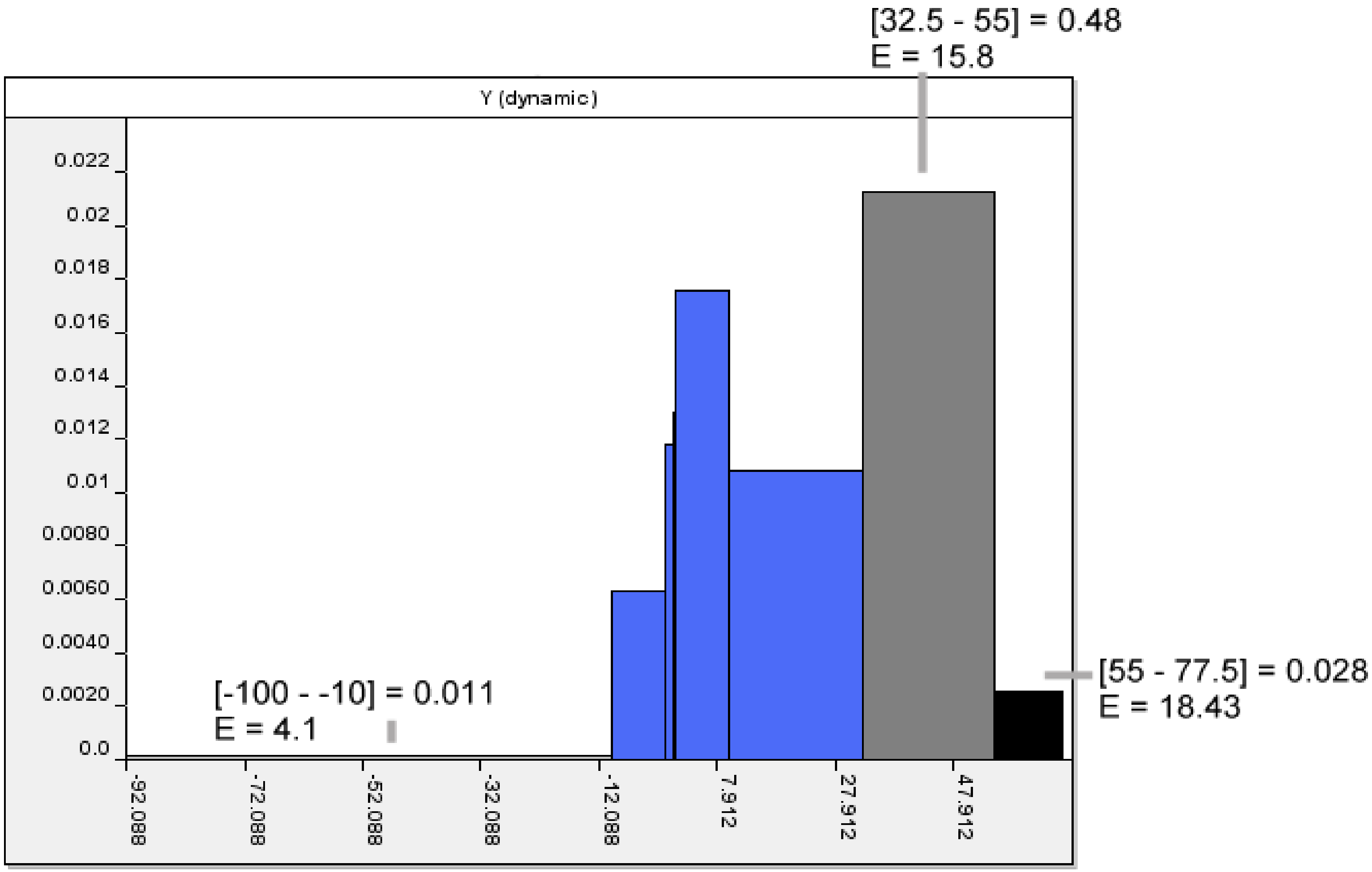
$$E\left[Y\right] = \int yf\left(y\right)dy = \lambda_1 \int yN\left(y \mid \mu_1, \sigma_1^2\right)dy + \lambda_2 \int yN\left(y \mid \mu_2, \sigma_2^2\right)dy = \lambda_1\mu_1 + \lambda_2\mu_2 = 30$$

$$Var\left[Y\right] = \lambda_1\left(\sigma_1^2 + \mu_1^2\right) + \lambda_2\left(\sigma_2^2 + \mu_2^2\right) - \left(\lambda_1\mu_1 + \lambda_2\mu_2\right)^2 = 455$$

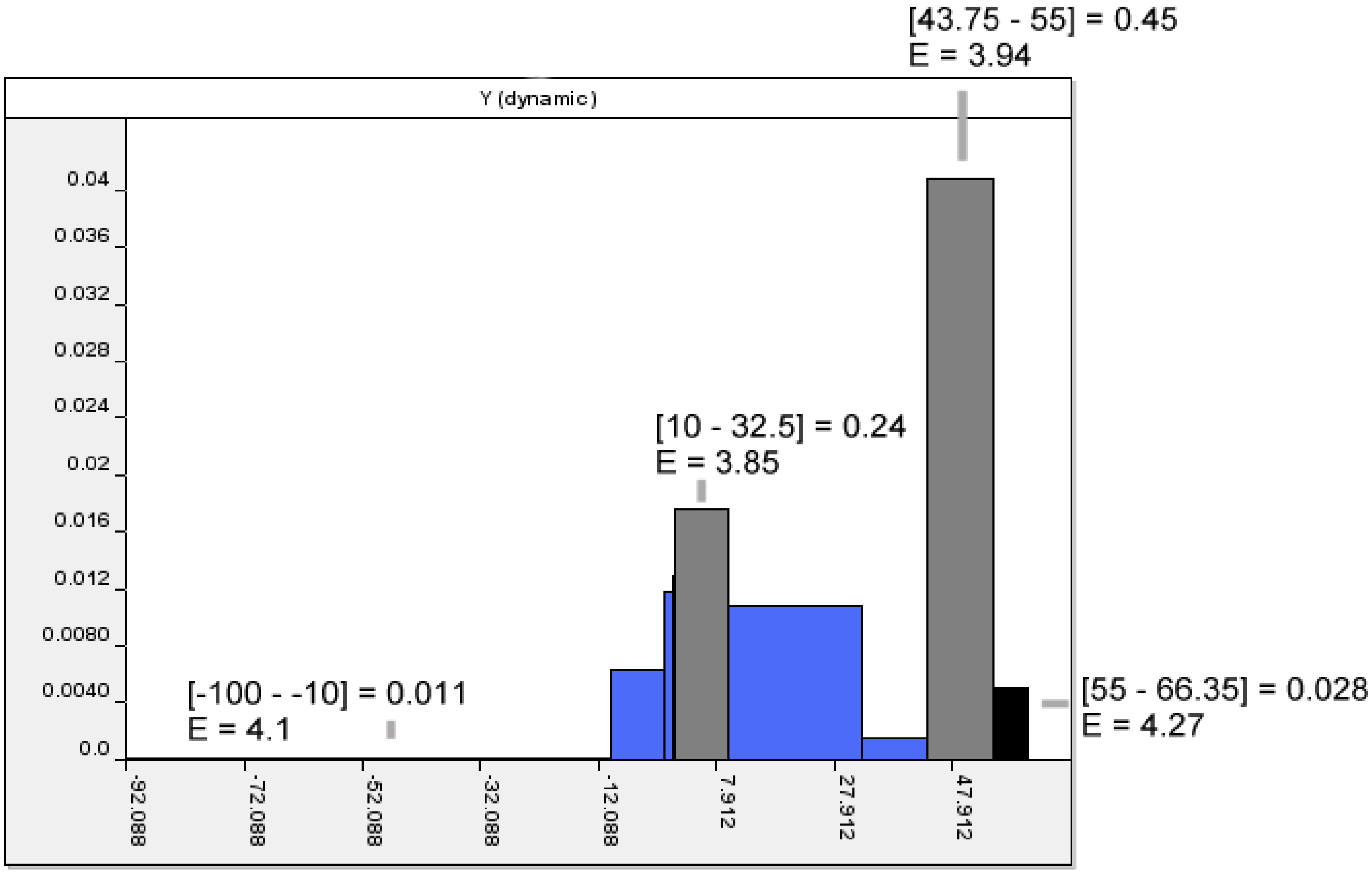
After 2 iterations



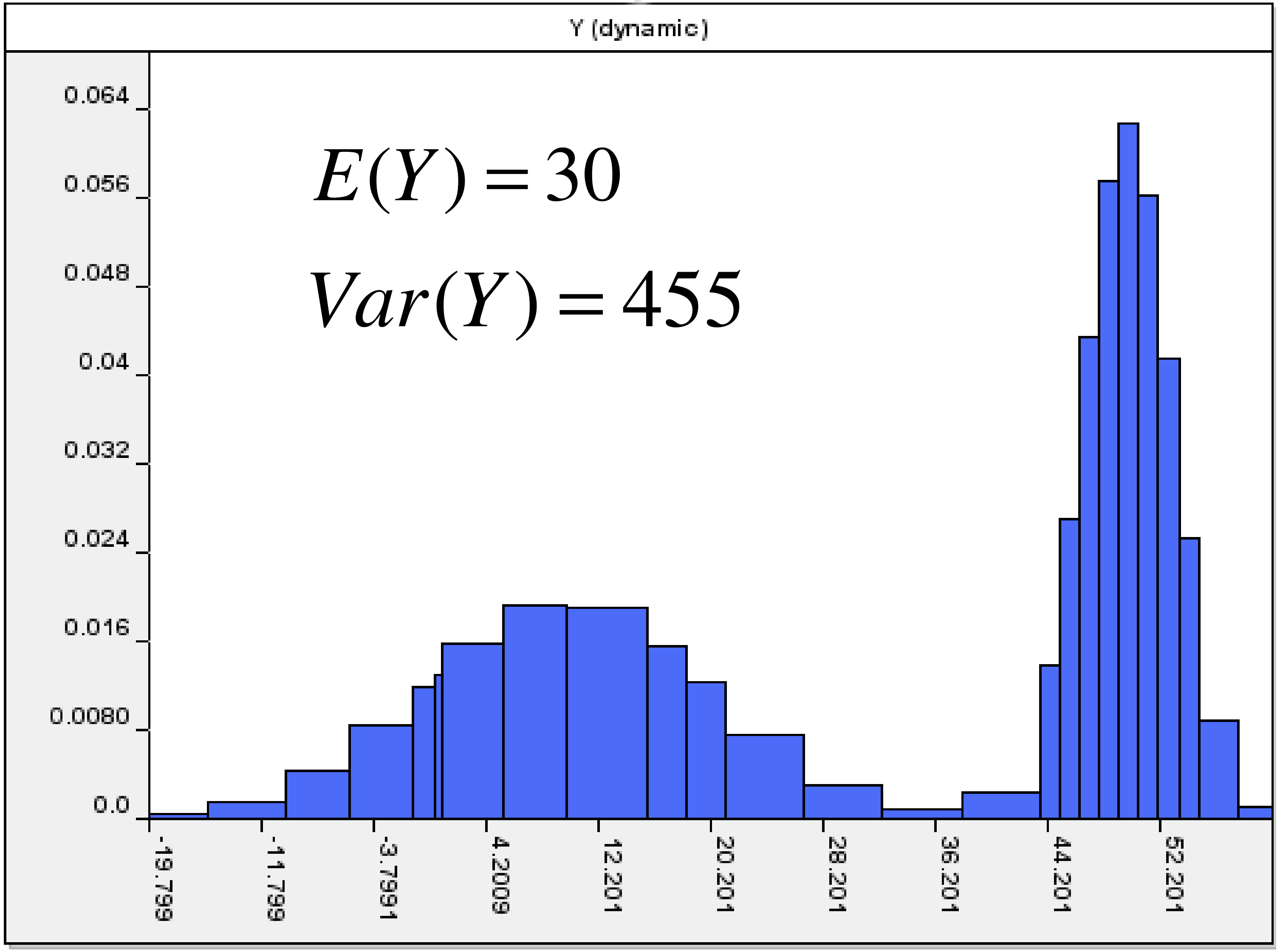
After 4 iterations



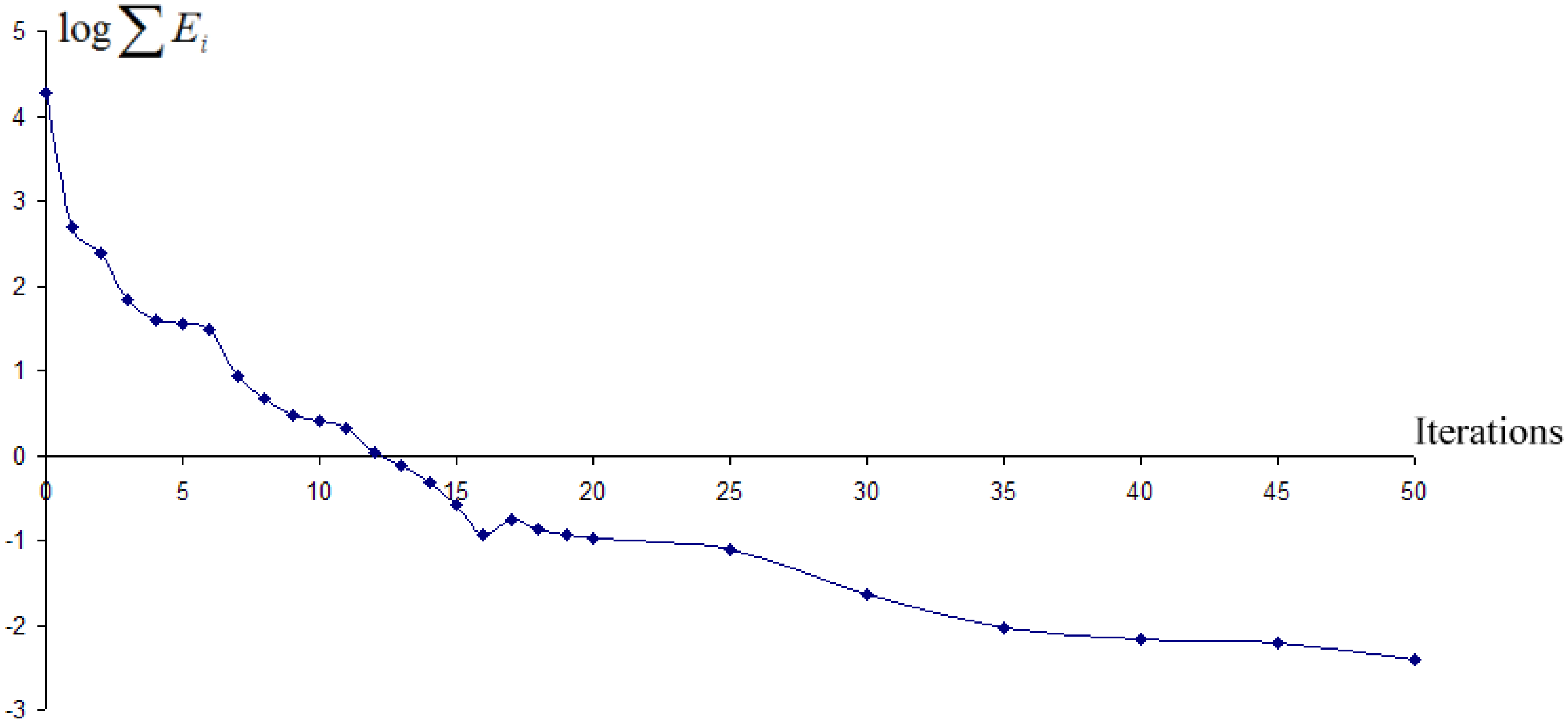
After 6 iterations



After 25 iterations



Convergence of sum Entropy Errors for Y

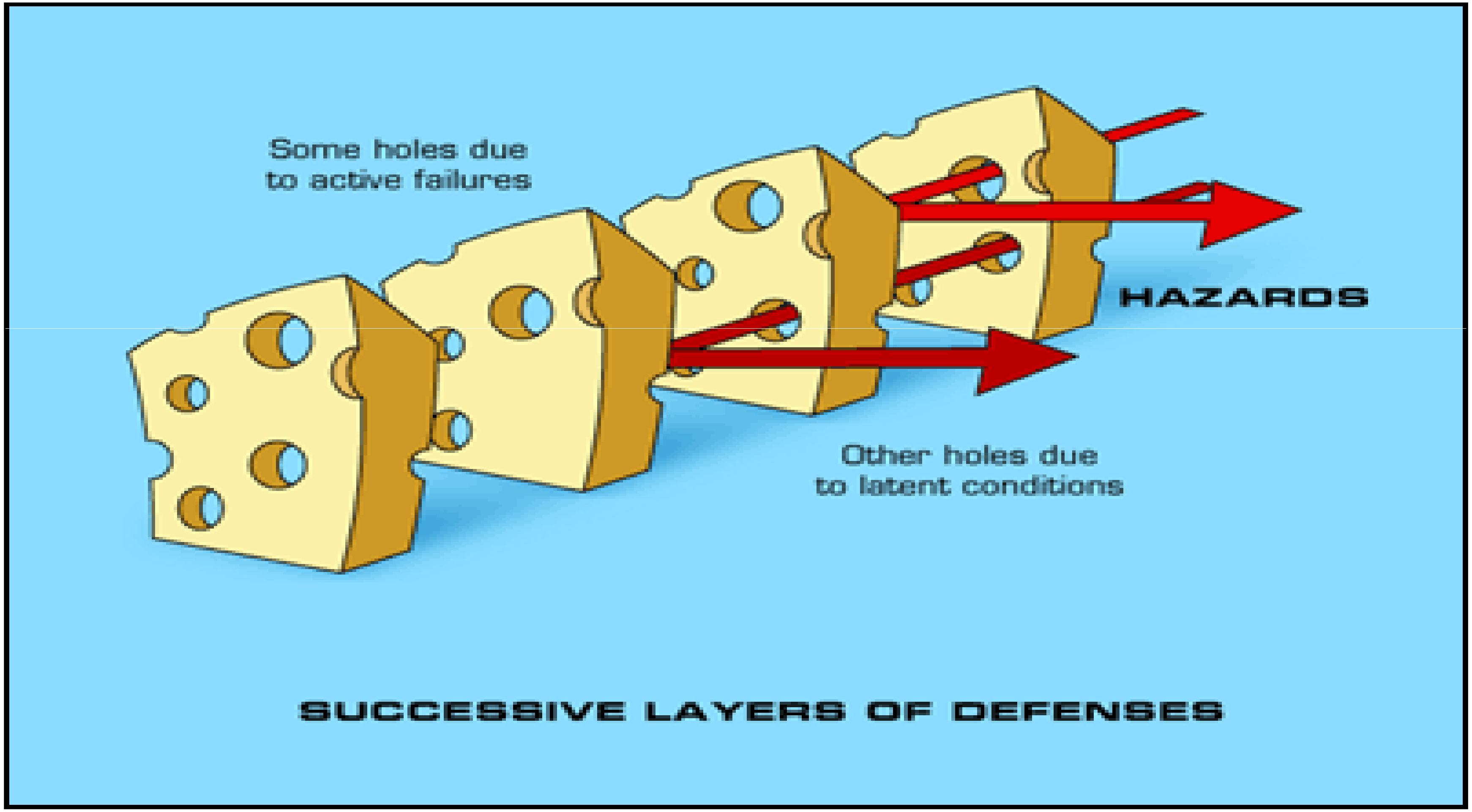


Operational risk

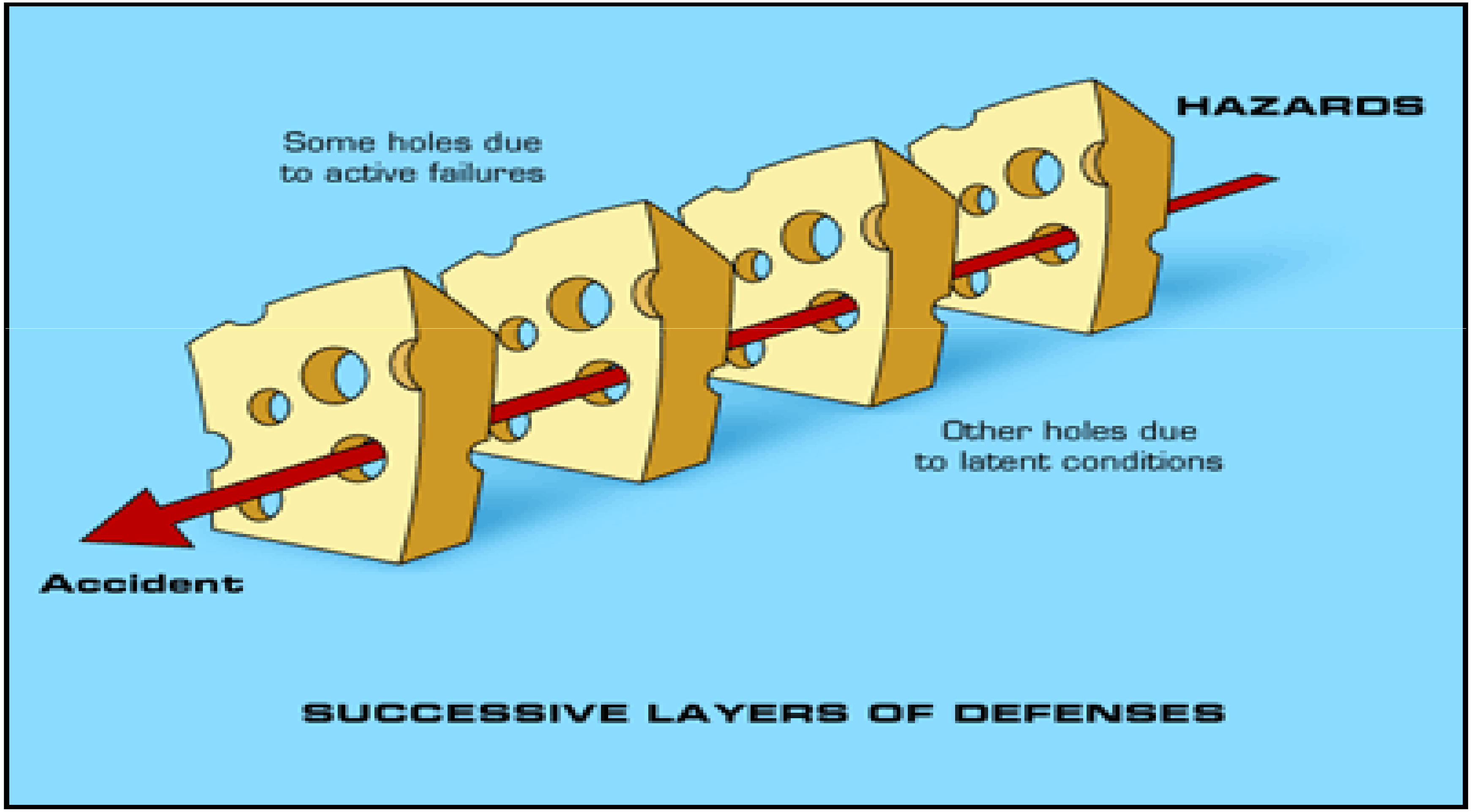
Resiliency Perspective

- Operational risk is faced by all organisations.
  - Human error main cause of a catastrophic event
  - ....but without latent weaknesses in organisation the event would not reach catastrophic proportions.
- OpRisk modelling cannot solely involve the investigation of statistical phenomena

The "Swiss Cheese" Accident Model



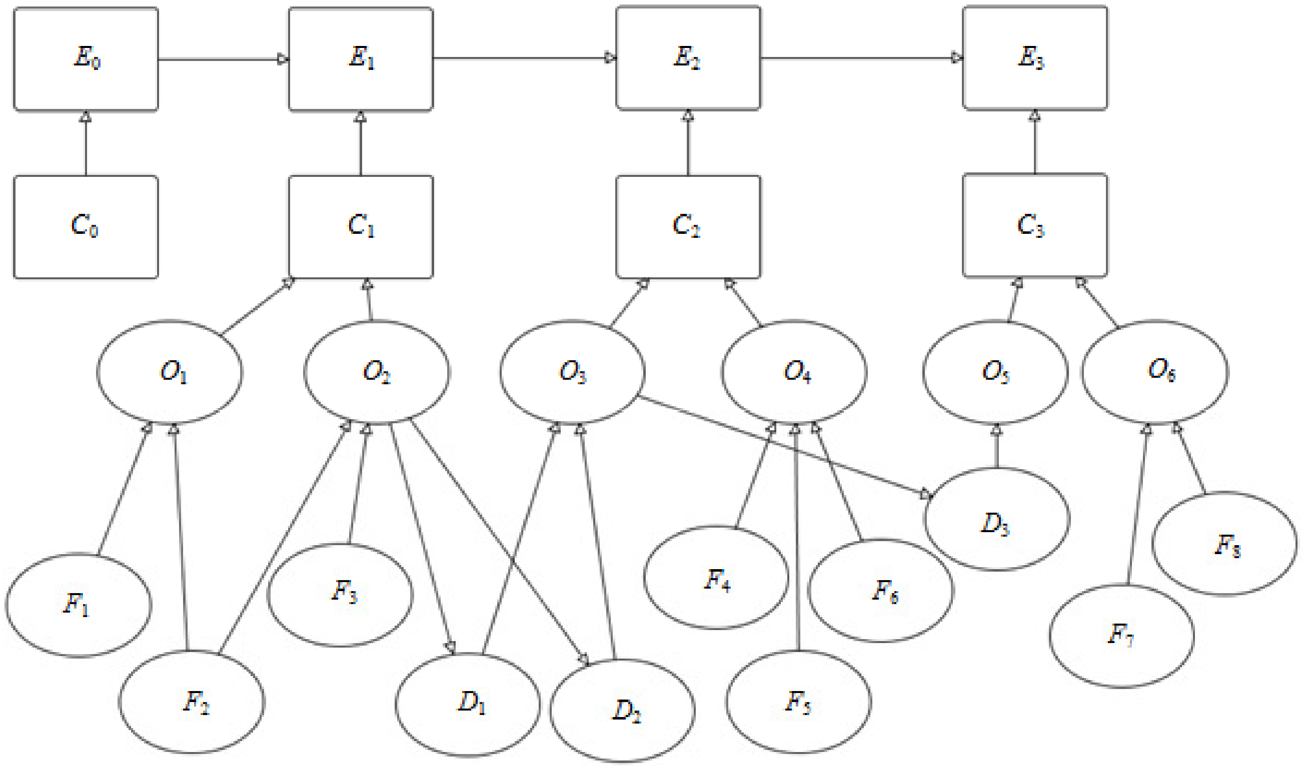
The "Swiss Cheese" Accident Model



Rogue Trading

- Process
  - Trade request – Conduct Trade – Registration of Trade – Reconciliation check, Settlement and Netting – Entering of Trade on Trading Books
- Controls (prevent, limit and mitigate loss)
  - Front office control environment (the control environment affects the probability of unauthorised trading)
  - Back office reconciliation checks (performed per trade)
  - Market positions and results monitoring, VaR calculation (periodical)
  - Audit checks (periodical but not as often as the market checks)

Loss Model



E - Events

C - Controls

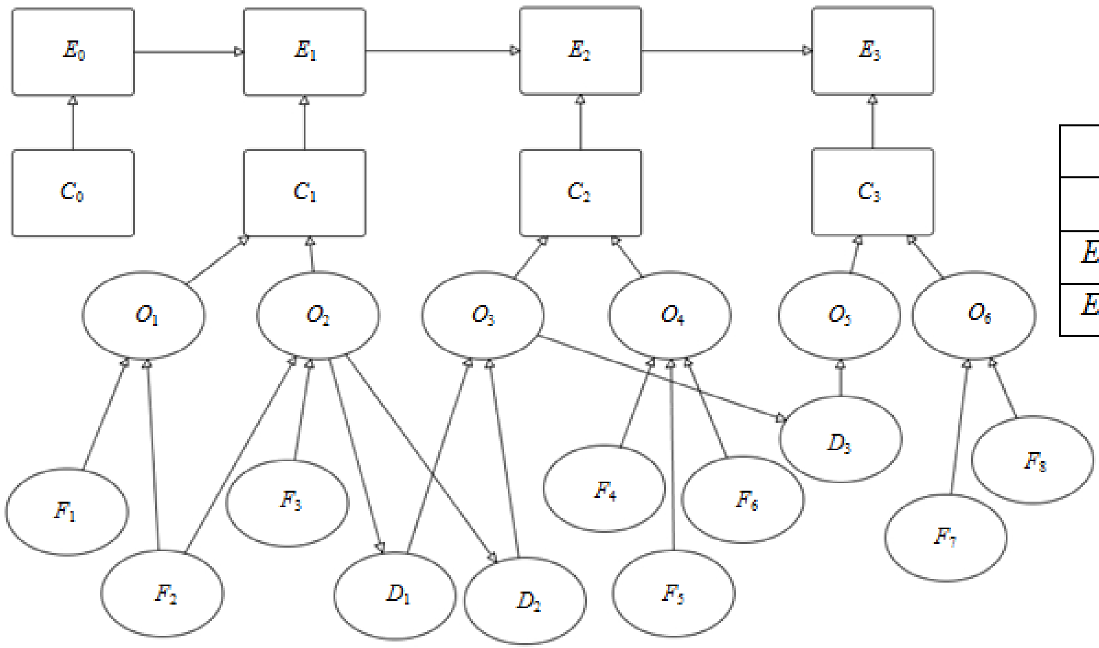
O – Operational Failures

C - Causal Factors

D – Dependency Factors

$$p(\mathrm{E},\mathrm{C},\mathrm{O},\mathrm{F},\mathrm{D}) = \prod_{t=1}^{T}\prod_{s=1}^{t}\prod_{j=1}^{m}\prod_{i=1}^{n}\prod_{k=1}^{o} p(E_t \mid E_{t-1}, C_t)\, p(C_t \mid \mathrm{O}_{C_t})\, p(O_j \mid \mathrm{F}_{O_j}, \mathrm{D}_{O_j})\, p(D_k \mid \mathrm{O}_{C_{t-s}})\, p(F_i)\, p(C_0)$$

Conditional Probability Tables



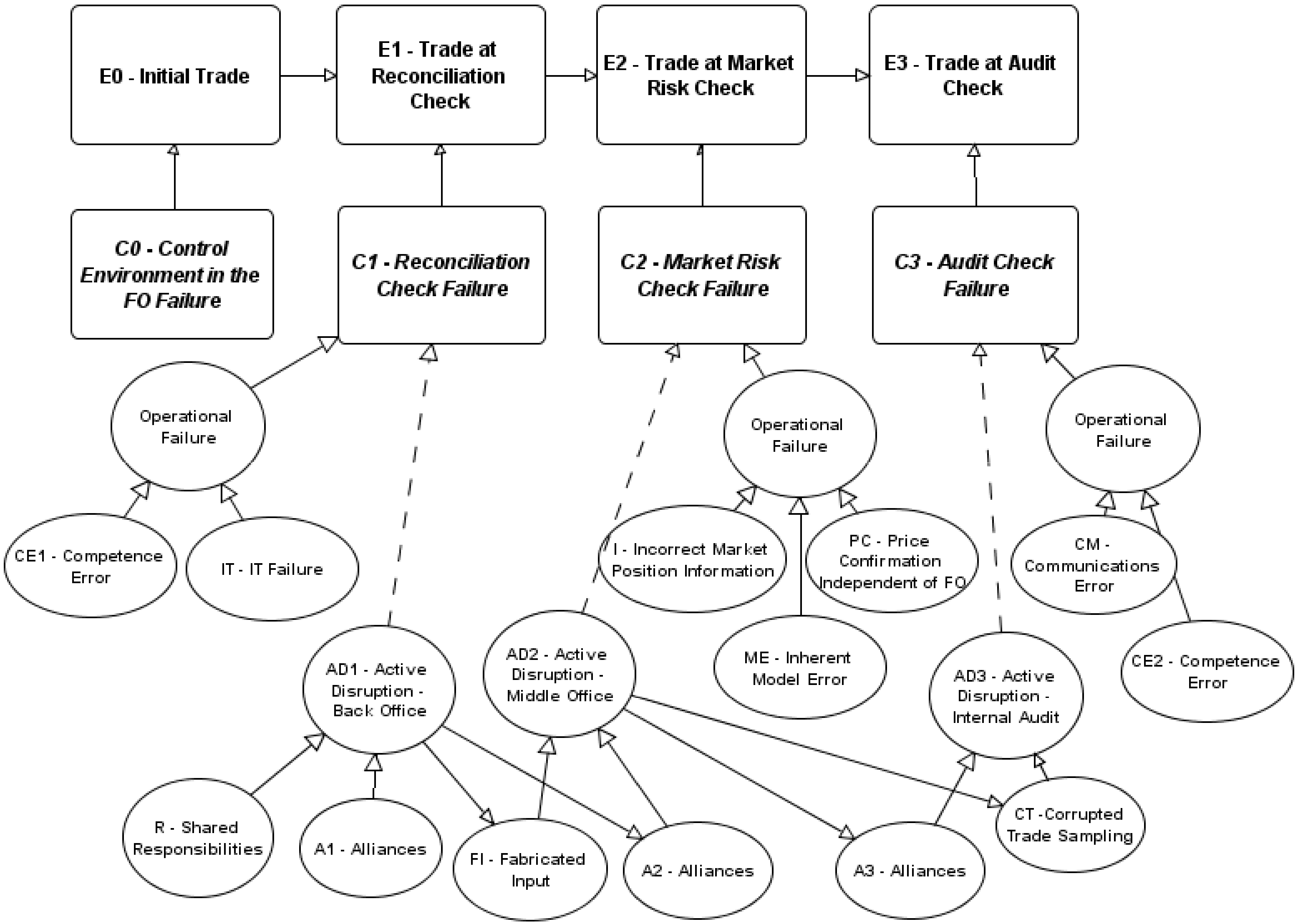
$$p(E_t \mid E_{t-1}, C_t)$$

	$E_{t-1} = fail$		$E_{t-1} = Ok$	
	$C_t = fail$	$C_t = Ok$	$C_t = fail$	$C_t = Ok$
$E_t = fail$	1	0	0	0
$E_t = Ok$	0	1	1	1

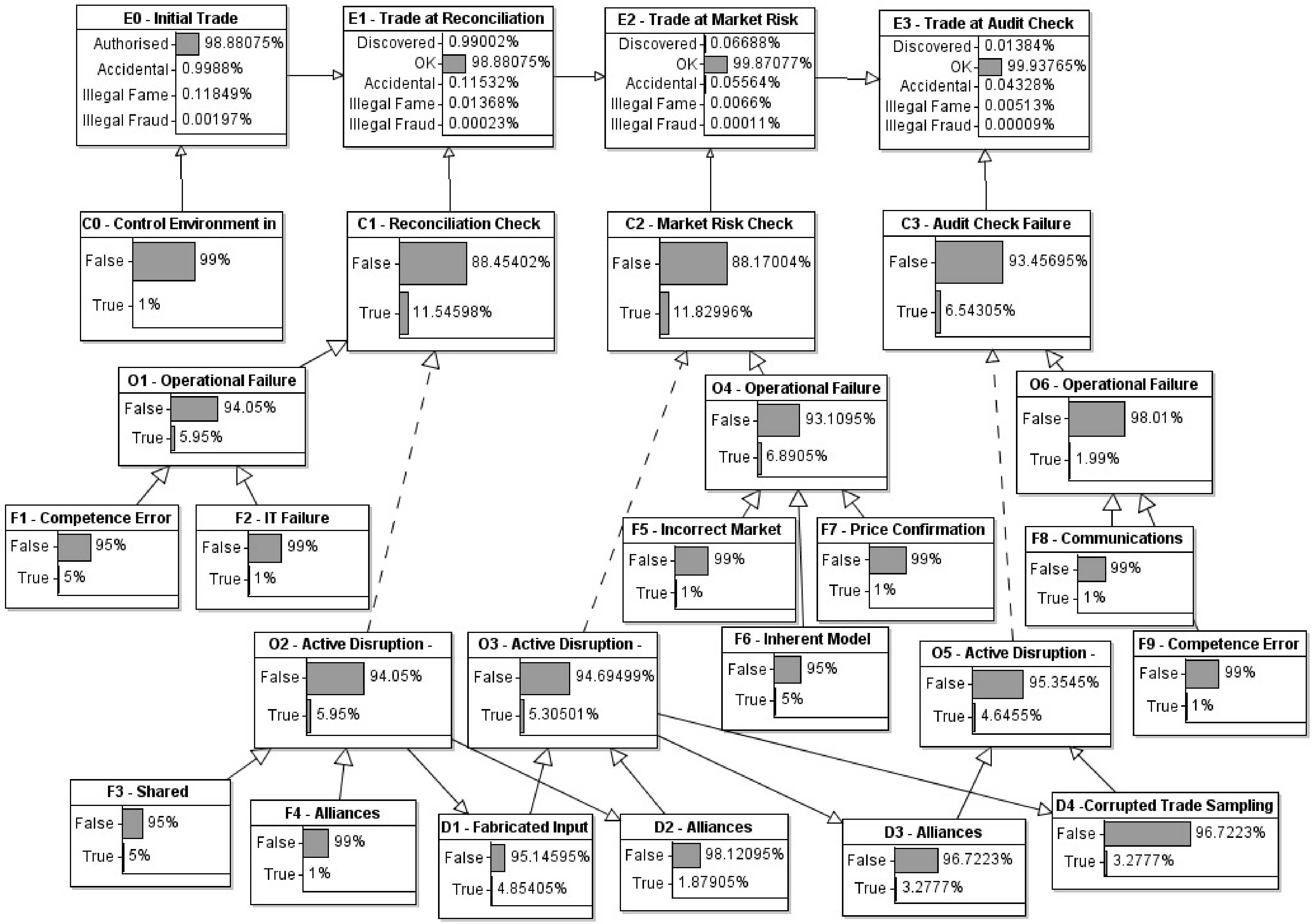
$$p(C_1 = fail \mid O_1, O_2) = \begin{cases} 1 & \text{if } O_1 \cup O_2 = fail \\ 0 & \text{otherwise} \end{cases}$$

	$O_1 = fail$	$O_1 = Ok$
$C_1 = fail$	0.8	0
$C_1 = Ok$	0.2	1

Example Loss Model



Executing the Loss Model



Loss Severity Model

$$p(E_t = e_t^L) = p_{e_t^L}$$

Probability of loss for particular event type

$$N_{e_t^L} \sim Bin(W_{e_{t-1}^L}, p_{e_t^L})$$

Number of events per type

$$W_{e_{t-1}^L} = V - \sum_{t=1}^{t-1} N_{e_{t-1}^L}$$

Volume of events (transactions)

$$T_{e_t^L} = S_{e_t^L} \times N_{e_t^L}$$

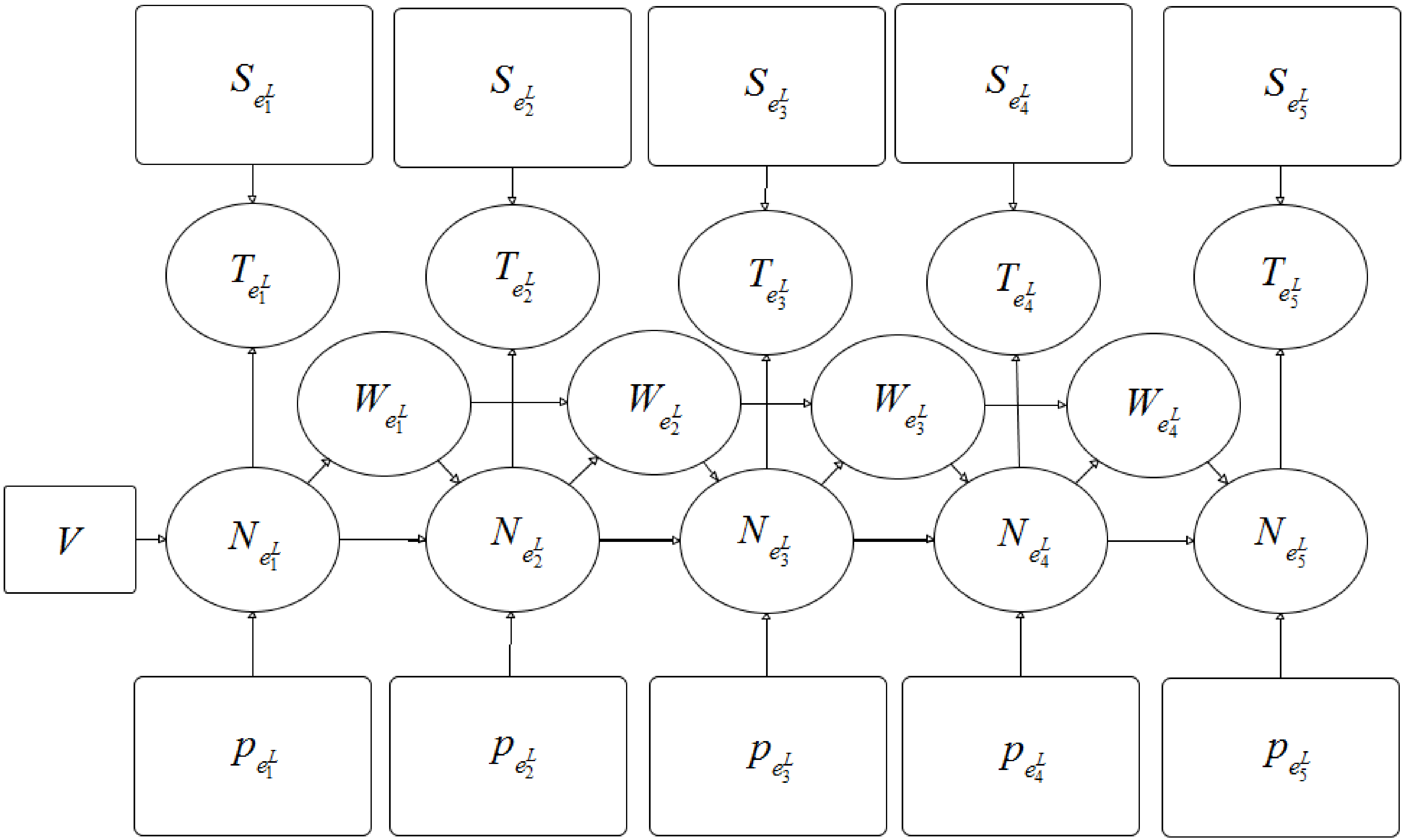
Total losses = severity x number of events

$$f(T_{e_t^L}) = \sum_{S_{e_t^L}, N_{e_t^L}, W_{e_{t-1}^L}} f(T_{e_t^L} \mid S_{e_t^L}, N_{e_t^L}) f(N_{e_t^L} \mid W_{e_{t-1}^L}, p_{e_t^L}) f(S_{e_t^L})$$

$$f(A) = \sum_{\mathbf{T}_{e_t^L}} f(A \mid \mathbf{T}_{e_t^L}) f(\mathbf{T}_{e_t^L})$$

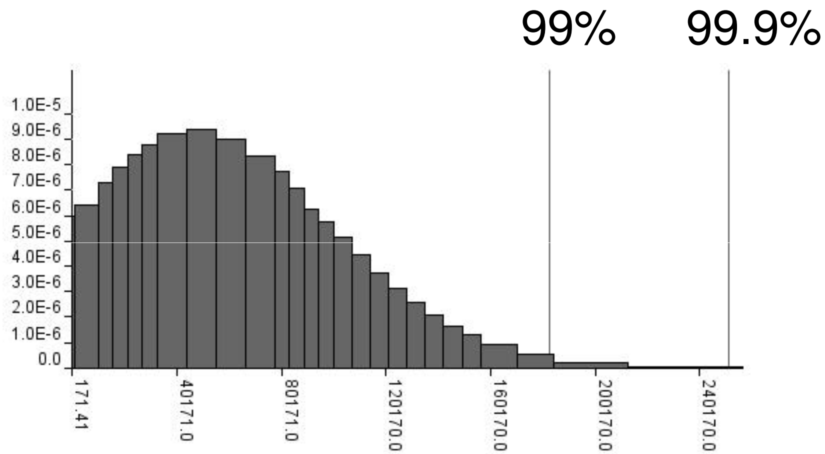
Total aggregated losses

Loss Severity Model



Executing Loss Severity Model

Loss severity	μ	σ^2
$S_{e_1^{discovered}}$	100	100,000
$S_{e_2^{discovered}}$	200	100,000
$S_{e_3^{discovered}}$	500	100,000
$S_{e_1^{accidental}}$	800	1,000,000
$S_{e_1^{illegalfame}}$	1000	1,000,000
$S_{e_1^{illegalfaud}}$	5000	1,000,000



Total losses for all events

Final Remarks

- Structured Method
  - Bayesian networks are extremely flexible
  - Can combine subjective judgements with data
  - Supports stress testing and scenarios
- Benefits for Operational Risk
  - Supports LDA approach
  - Supports causal modelling
  - Flexible modelling framework
  - Optimise capital allocation

Some References

Finance

- Neil, M. and Hager, D., **Modeling Operational Risk in Financial Institutions using Hybrid Dynamic Bayesian Networks**. Journal of Operational Risk, Volume 4/Number 1, Spring, 2009.
- Neil M. Marquez D. And Fenton N. **Using Bayesian Networks to model operational risk to information technology infrastructure in financial institutions**. Cass-Capco Institute Paper series on risk - Journal Vol. 22., April 2008.

Engineering

- Neil M. Marquez D. Fenton N. **Improved Reliability Modeling using Bayesian Networks and Dynamic Discretization**. Reliability Engineering & System Safety, Volume 95, Issue 4, April 2010, Pages 412-425.

Algorithm

- Neil M. Tailor M. Marquez D**. Inference in Bayesian Networks using Dynamic Discretisation**. Statistics and Computing 17:3 September 2007.