The BDe score

A brief derivation

Dirk Husmeier

The parent node set of node X_n in $\mathcal{G}$, $\pi_n = \pi_n(\mathcal{G})$, is the set of all nodes from which an edge points to node X_n in $\mathcal{G}$. The conditional dependence of X_n on π_n is described by the conditional probability (density) $P(X_n|\pi_n, \boldsymbol{\theta}_{n,\pi_n})$, which lies in a chosen function family, and depends on the parameters $\boldsymbol{\theta}_{n,\pi_n}$. Consider a set of experimental conditions $t, 1 \leq t \leq M$. Let $X_n(t)$ represent the realization of X_n in experimental condition t. We are given a data set $\mathcal{D}$, where $\mathcal{D}_{n,t}$ and $\mathcal{D}_{(\pi_n,t)}$ are the tth realizations $X_n(t)$ and $\pi_n(t)$ of X_n and π_n, respectively.

$$P(\mathcal{D}|\mathcal{G}) = \int P(\mathcal{D}|\mathcal{G}, \boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathcal{G}) d\boldsymbol{\theta}$$

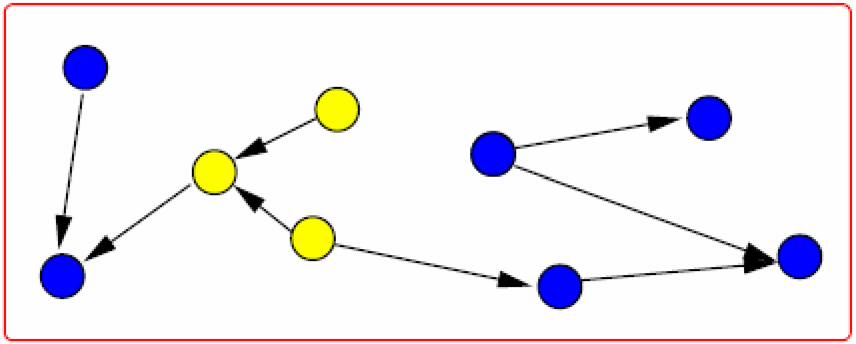
$$P(\mathcal{D}|\mathcal{G}, \boldsymbol{\theta}) = \prod_{n=1}^{N} \prod_{t=1}^{M} P\Big(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big)$$

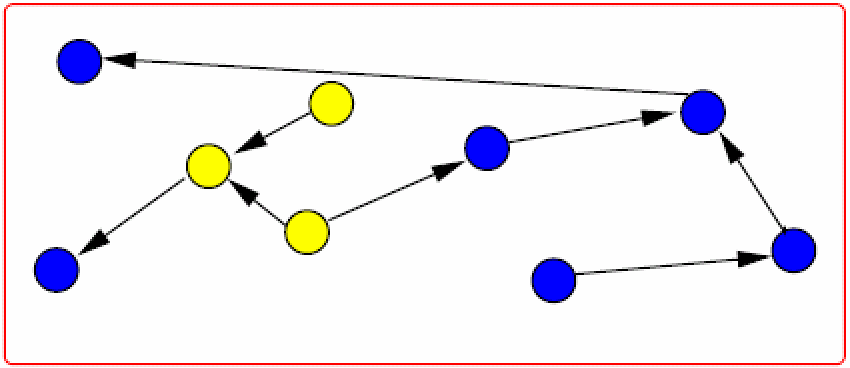
Parameter modularity:

$$P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}) \;=\; P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}') \;:=\; P(\boldsymbol{\theta}_{n,\pi_n})$$

Parameter independence:

$$P(\boldsymbol{\theta}|\mathcal{G}) \;=\; \prod_{n=1}^{N} P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G})$$





$$P(\mathcal{D}|\mathcal{G}) \;=\; \int P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta}$$

$$P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta}) \;=\; \prod_{n=1}^{N}\prod_{t=1}^{M} P\Big(X_n(t)=\mathcal{D}_{n,t}|\pi_n(t)=\mathcal{D}_{(\pi_n,t)},\boldsymbol{\theta}_{n,\pi_n}\Big)$$

Parameter modularity:

$$P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}) \;=\; P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}') \;:=\; P(\boldsymbol{\theta}_{n,\pi_n})$$

Parameter independence:

$$P(\boldsymbol{\theta}|\mathcal{G}) \;=\; \prod_{n=1}^{N} P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G})$$

Decomposition of the marginal likelihood

$$
\begin{aligned}
P(\mathcal{D}|\mathcal{G}) &= \int P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta} \\
&= \int \prod_{n=1}^{N}\prod_{t=1}^{M} P\Big(X_n(t)=\mathcal{D}_{n,t}|\pi_n(t)=\mathcal{D}_{(\pi_n,t)},\boldsymbol{\theta}_{n,\pi_n}\Big)P(\boldsymbol{\theta}_{n,\pi_n})d\boldsymbol{\theta}_{n,\pi_n}
\end{aligned}
$$

$$
\begin{aligned}
P(\mathcal{D}|\mathcal{G}) &= \int P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta} \;=\; \prod_{n=1}^{N}\Psi(\mathcal{D}_n^{\pi_n}) \\
\Psi(\mathcal{D}_n^{\pi_n}) &= \int \prod_{t=1}^{M} P\Big(X_n(t)=\mathcal{D}_{n,t}|\pi_n(t)=\mathcal{D}_{(\pi_n,t)},\boldsymbol{\theta}_{n,\pi_n}\Big)P(\boldsymbol{\theta}_{n,\pi_n})d\boldsymbol{\theta}_{n,\pi_n}
\end{aligned}
$$

where $\mathcal{D}_n^{\pi_n} := \{(\mathcal{D}_{n,t},\mathcal{D}_{\pi_n,t}) : 1 \le t \le m\}$

Decomposition of the marginal likelihood

$$P(\mathcal{D}|\mathcal{G}) = \int P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta}$$
$$= \int \prod_{n=1}^{N}\prod_{t=1}^{M} P\Big(X_n(t) = \mathcal{D}_{n,t}|\pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big)P(\boldsymbol{\theta}_{n,\pi_n})d\boldsymbol{\theta}_{n,\pi_n}$$

$$P(\mathcal{D}|\mathcal{G}) = \int P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta} = \prod_{n=1}^{N}\Psi(\mathcal{D}_n^{\pi_n})$$

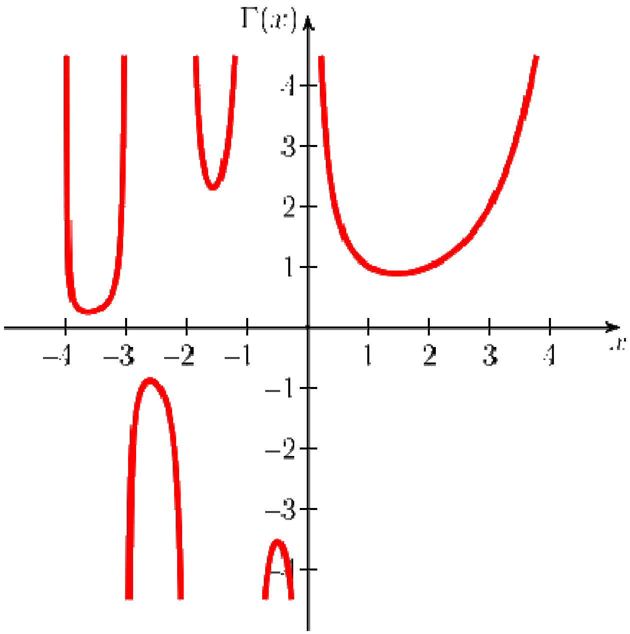
$$\Psi(\mathcal{D}_n^{\pi_n}) = \int \prod_{t=1}^{M} P\Big(X_n(t) = \mathcal{D}_{n,t}|\pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big)P(\boldsymbol{\theta}_{n,\pi_n})d\boldsymbol{\theta}_{n,\pi_n}$$

where $\mathcal{D}_n^{\pi_n} := \{(\mathcal{D}_{n,t}, \mathcal{D}_{\pi_n,t}) : 1 \leq t \leq m\}$

Analytically tractable?

Mathematical Primer

$$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)dt$$

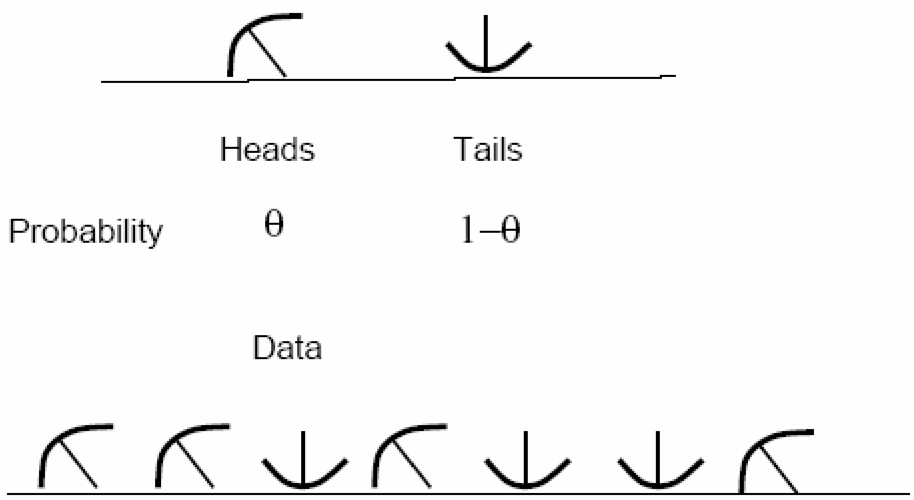


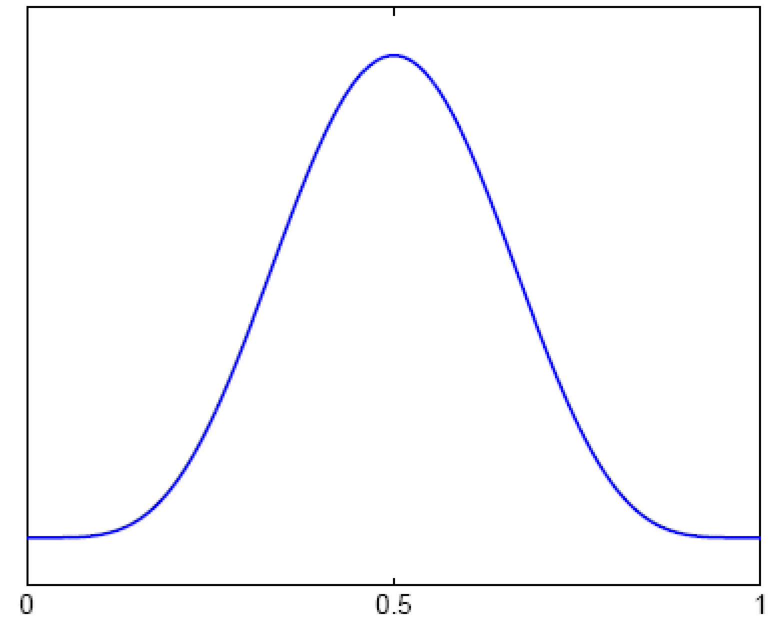
$$\begin{aligned} \Gamma(z+1) &= z\Gamma(z) \\ \Gamma(1) &= 1 \\ \Gamma(n+1) &= n! \end{aligned}$$

$$\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \mathcal{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Binomial distribution

$$L(\theta) \;=\; P(\mathcal{D}|\theta)$$





Total number of observations

Likelihood:

$$L(\theta) \;=\; P(\mathcal{D}|\theta) \;=\; \binom{M}{m}\theta^m(1-\theta)^{M-m}$$

Number of heads = sufficient statistics

Binomial distribution

Tossing a coin, $\theta \in [0,1]$, M observations, m times heads.

Likelihood:

$$L(\theta) \;=\; P(\mathcal{D}|\theta) \;=\; \binom{M}{m}\theta^m(1-\theta)^{M-m}$$

Log Likelihood:

$$\log L(\theta) \;=\; \log P(\mathcal{D}|\theta) \;=\; m\log\theta + (M-m)\log(1-\theta) + C$$

Maximum Likelihood:

$$\begin{aligned}\frac{d\log L(\theta)}{d\theta} &= \frac{m}{\theta} - \frac{M-m}{1-\theta} = 0\\ m(1-\theta) &= (M-m)\theta\\ \theta &= \frac{m}{M}\end{aligned}$$

Tossing a coin, $\theta \in [0, 1]$, M observations, m times heads.

Likelihood:

$$L(\theta) = P(\mathcal{D}|\theta) = \binom{M}{m} \theta^m (1-\theta)^{M-m}$$

Conjugate Prior:

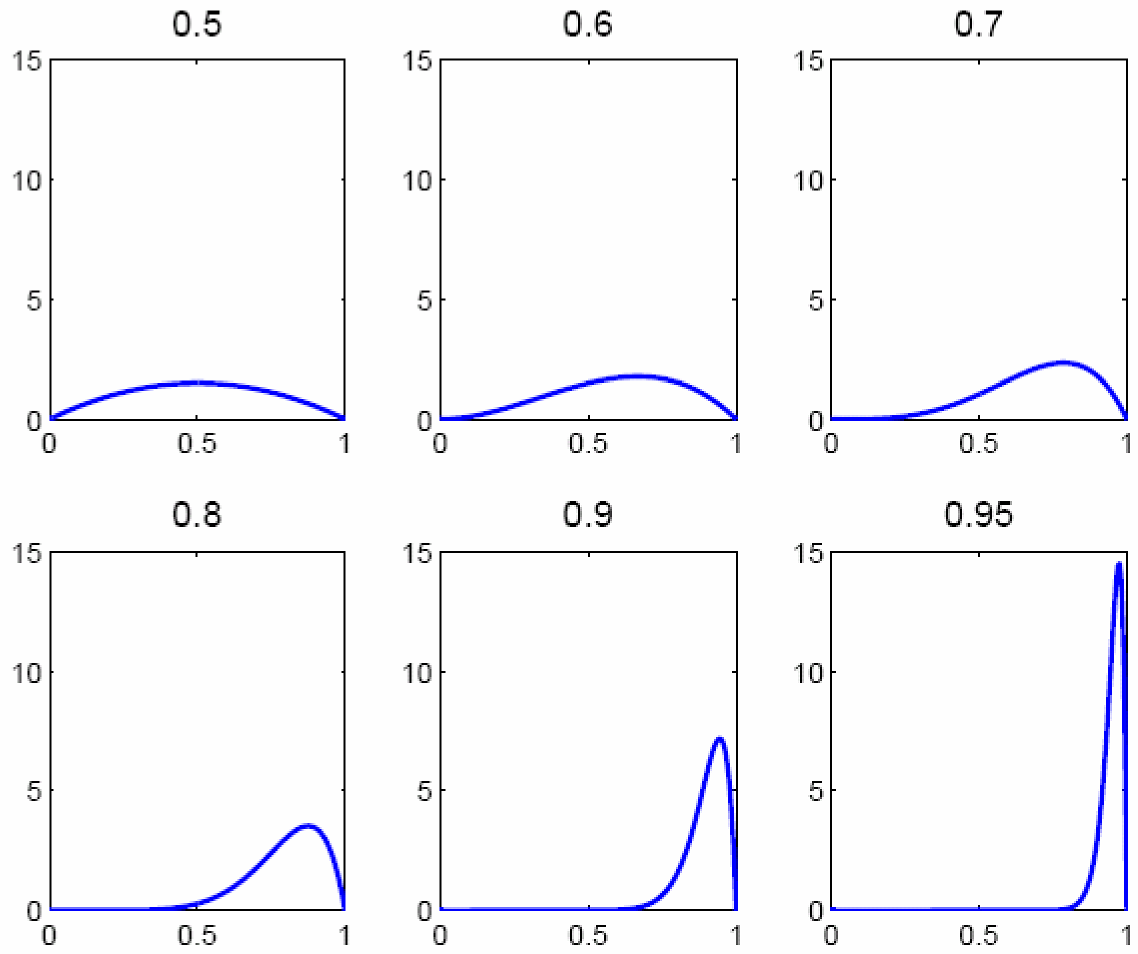
$$P(\theta) = \frac{1}{Z} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

Posterior:

$$\begin{aligned}
P(\theta|\mathcal{D}) &\propto P(\mathcal{D}|\theta) P(\theta) \\
&\propto \theta^m (1-\theta)^{M-m} \times \theta^{\alpha-1} (1-\theta)^{\beta-1} \\
&= \theta^{(m+\alpha)-1} (1-\theta)^{(\beta+M-m)-1} \\
&= \theta^{\tilde{\alpha}-1} (1-\theta)^{\tilde{\beta}-1}
\end{aligned}$$

$$\tilde{\alpha} = \alpha + m, \;\; \tilde{\beta} = \beta + M - m$$

Beta Prior, $\beta = 2$, $\mu = \alpha/(\alpha + \beta)$



α=β=1 → uniform prior over [0,1]

Conjugate Prior:

$$P(\theta) = \frac{1}{Z}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Normalization constant:

$$Z = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \mathcal{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)dt$$

$$\begin{aligned}\Gamma(z+1) &= z\Gamma(z) \\ \Gamma(1) &= 1 \\ \Gamma(n+1) &= n!\end{aligned}$$

Prior:

$$P(\theta) \;=\; \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Posterior:

$$P(\theta|\mathcal{D}) \;=\; \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\theta^{\tilde{\alpha}-1}(1-\theta)^{\tilde{\beta}-1}$$

$\tilde{\alpha} = \alpha + m, \;\; \tilde{\beta} = \beta + M - m$

Posterior:

$$P(\theta|\mathcal{D}) = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\theta^{\tilde{\alpha}-1}(1-\theta)^{\tilde{\beta}-1}$$

Posterior mean:

$$\begin{aligned}\langle\theta\rangle &= \int_0^1 \theta\, P(\theta|\mathcal{D})\, d\theta \\ &= \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\int_0^1 \theta^{\tilde{\alpha}+1-1}(1-\theta)^{\tilde{\beta}-1}\, d\theta\end{aligned}$$

Posterior:

$$P(\theta|\mathcal{D}) = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\theta^{\tilde{\alpha}-1}(1-\theta)^{\tilde{\beta}-1}$$

Recall: $$\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \Gamma(z+1) = z\Gamma(z)$$

Posterior mean:

$$\begin{aligned}
\langle\theta\rangle &= \int_0^1 \theta\, P(\theta|\mathcal{D})\, d\theta \\
&= \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \int_0^1 \theta^{\tilde{\alpha}+1-1}(1-\theta)^{\tilde{\beta}-1}\, d\theta \\
&= \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \times \frac{\Gamma(\tilde{\alpha}+1)\Gamma(\tilde{\beta})}{\Gamma(\tilde{\alpha}+1+\tilde{\beta})} \quad = \quad \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \times \frac{\tilde{\alpha}\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}{(\tilde{\alpha}+\tilde{\beta})\Gamma(\tilde{\alpha}+\tilde{\beta})}
\end{aligned}$$

Posterior:

$$P(\theta|\mathcal{D}) = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\theta^{\tilde{\alpha}-1}(1-\theta)^{\tilde{\beta}-1}$$

Recall:

$$\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \Gamma(z+1) = z\Gamma(z)$$

Posterior mean:

$$\begin{aligned}
\langle\theta\rangle &= \int_0^1 \theta\, P(\theta|\mathcal{D})\, d\theta \\
&= \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\int_0^1 \theta^{\tilde{\alpha}+1-1}(1-\theta)^{\tilde{\beta}-1}\, d\theta \\
&= \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \times \frac{\Gamma(\tilde{\alpha}+1)\Gamma(\tilde{\beta})}{\Gamma(\tilde{\alpha}+1+\tilde{\beta})} = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \times \frac{\tilde{\alpha}\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}{(\tilde{\alpha}+\tilde{\beta})\Gamma(\tilde{\alpha}+\tilde{\beta})} \\
&= \frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} = \frac{m+\alpha}{[m+\alpha]\,+\,(M-m+\beta)} = \frac{m+\alpha}{M+\alpha+\beta}
\end{aligned}$$

Binomial distribution

Tossing a coin, $\theta \in [0, 1]$, M observations, m times heads.

Likelihood:

$$L(\theta) \;=\; P(\mathcal{D}|\theta) \;=\; \binom{M}{m} \theta^m (1-\theta)^{M-m}$$

Multinomial distribution

K outcomes, parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k, \ldots, \theta_K) \in \mathcal{S}$, i.e. $\theta_k \geq 0 \; \forall k$, $\sum_{k=1}^{K} \theta_k = 1$. M observations, M_k in category k, $k \in \{1, \ldots, K\}$.

Likelihood:

$$L(\theta) \;=\; P(\mathcal{D}|\theta) \;=\; \frac{M!}{\prod_k M_k!} \prod_{k=1}^{K} \theta^{M_k}$$

Multinomial distribution

K outcomes, parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k, \ldots, \theta_K) \in \mathcal{S}$, i.e. $\theta_k \geq 0 \; \forall k$, $\sum_{k=1}^{K} \theta_k = 1$. M observations, M_k in category k, $k \in \{1, \ldots, K\}$.

Likelihood:

$$L(\theta) \; = \; P(\mathcal{D}|\theta) \; = \; \frac{M!}{\prod_k M_k!} \prod_{k=1}^{K} \theta^{M_k}$$

Conjugate Prior:

$$P(\boldsymbol{\theta}) \; = \; Dir(\boldsymbol{\theta}) \; = \; \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Note:

$$\int_{\boldsymbol{\theta} \in \mathcal{S}} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} \; = \; \left(\frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \right)^{-1}$$

Posterior:

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}) = \prod_{k=1}^{K} \theta^{M_k} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} = \prod_{k=1}^{K} \theta_k^{M_k + \alpha_k - 1} = \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1}$$

Marginal likelihood:

$$P(\mathcal{D}|\boldsymbol{\alpha}) = \int_{\mathcal{S}} P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\alpha})d\boldsymbol{\theta} = \int_{\mathcal{S}} \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1} d\boldsymbol{\theta} = \frac{\prod_k \Gamma(\tilde{\alpha}_k)}{\Gamma(\sum_k \tilde{\alpha}_k)}$$

Posterior:

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\alpha})}{P(\mathcal{D}|\boldsymbol{\alpha})} = \frac{\Gamma(\sum_k \tilde{\alpha}_k)}{\prod_k \Gamma(\tilde{\alpha}_k)} \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1}$$

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{\Gamma(\sum_k \tilde{\alpha}_k)}{\prod_k \Gamma(\tilde{\alpha}_k)} \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1}$$

$$\int_{\boldsymbol{\theta}\in\mathcal{S}} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} = \left(\frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)}\right)$$

$$\Gamma(z+1) = z\Gamma(z)$$

Posterior mean:

$$\begin{aligned}
\langle\theta_1\rangle &= \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1 \, P(\boldsymbol{\theta}|\mathcal{D}) \, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1^{\tilde{\alpha}_1 + 1 - 1} \prod_{k=2}^{K} \theta_k^{\tilde{\alpha}_k - 1} \, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1 + 1) \prod_{k=2}^{K} \Gamma(\tilde{\alpha}_k)}{\Gamma(\tilde{\alpha}_1 + 1 + \sum_{k=2}^{K} \tilde{\alpha}_k)}
\end{aligned}$$

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{\Gamma(\sum_k \tilde{\alpha}_k)}{\prod_k \Gamma(\tilde{\alpha}_k)} \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1}$$

$$\int_{\boldsymbol{\theta}\in\mathcal{S}} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} = \left(\frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)}\right)$$

$$\Gamma(z+1) = z\Gamma(z)$$

Posterior mean:

$$\begin{aligned}
\langle\theta_1\rangle &= \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1\, P(\boldsymbol{\theta}|\mathcal{D})\, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1^{\tilde{\alpha}_1+1-1} \prod_{k=2}^{K} \theta_k^{\tilde{\alpha}_k-1}\, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1+1)\prod_{k=2}^{K}\Gamma(\tilde{\alpha}_k)}{\Gamma(\tilde{\alpha}_1+1+\sum_{k=2}^{K}\tilde{\alpha}_k)} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\Gamma(\tilde{\alpha}_1+1+\sum_{k=2}^{K}\tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1+1)\prod_{k=2}^{K}\Gamma(\tilde{\alpha}_k)}{\prod_{k=1}^{K}\Gamma(\tilde{\alpha}_k)} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\Gamma(1+\sum_{k=1}^{K}\tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1+1)}{\Gamma(\tilde{\alpha}_1)} = \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{(\sum_{k=1}^{K}\tilde{\alpha}_k)\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_k)} \times \frac{\tilde{\alpha}_1\Gamma(\tilde{\alpha}_1)}{\Gamma(\tilde{\alpha}_1)}
\end{aligned}$$

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{\Gamma(\sum_k \tilde{\alpha}_k)}{\prod_k \Gamma(\tilde{\alpha}_k)} \prod_{k=1}^{K} \theta_k^{\tilde{\alpha}_k - 1}$$

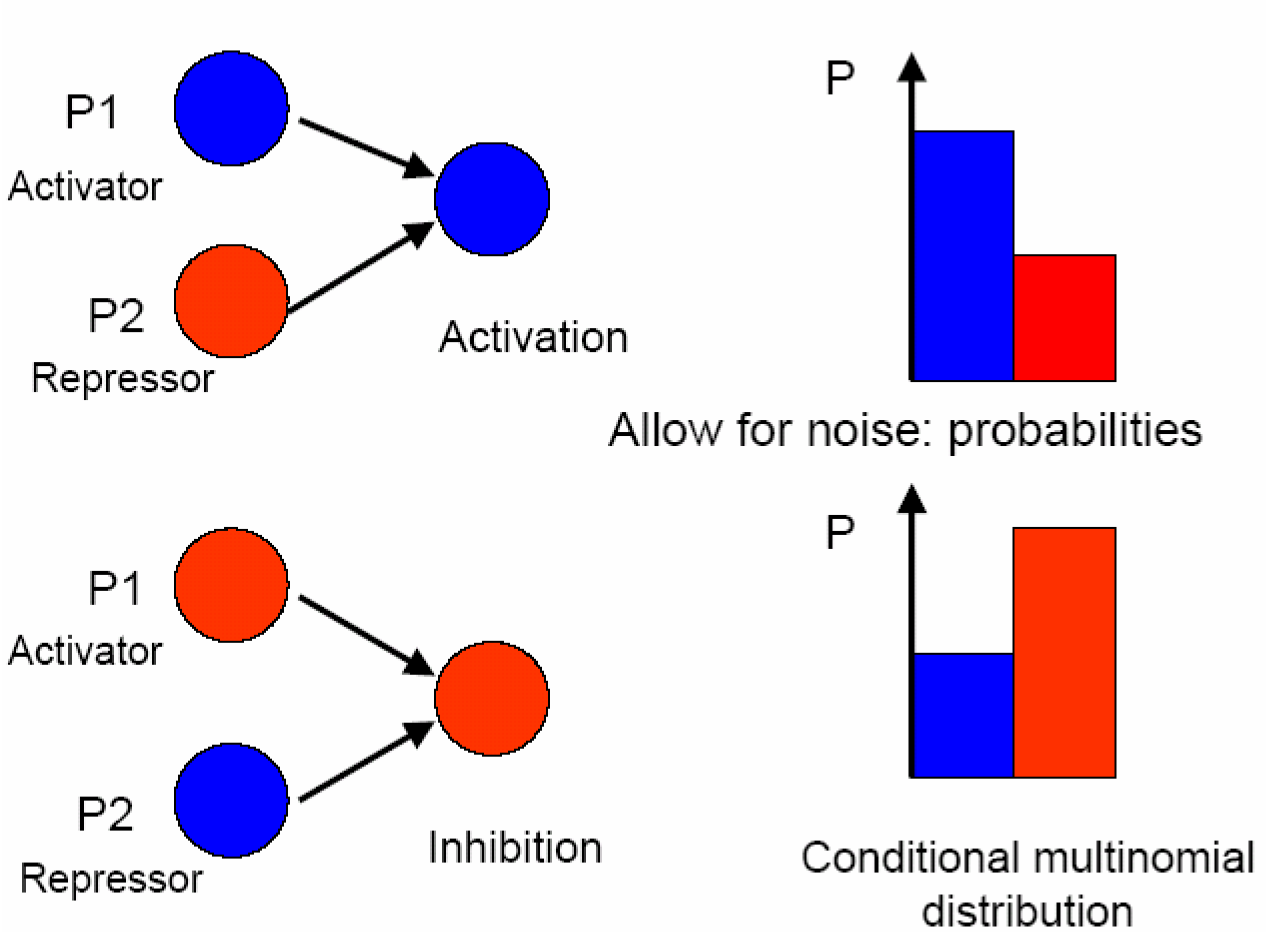
$$\int_{\boldsymbol{\theta}\in\mathcal{S}} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} = \left(\frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)}\right)$$

$$\Gamma(z+1) = z\Gamma(z)$$

Posterior mean:

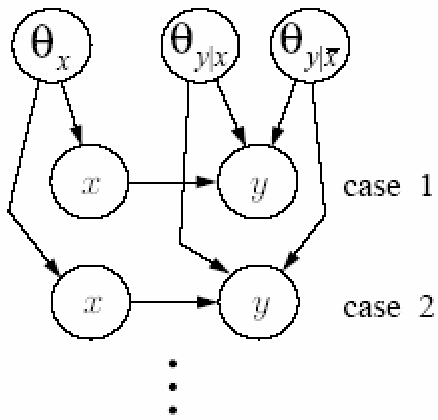
$$\begin{aligned}
\langle\theta_1\rangle &= \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1 \, P(\boldsymbol{\theta}|\mathcal{D}) \, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \int_{\boldsymbol{\theta}\in\mathcal{S}} \theta_1^{\tilde{\alpha}_1 + 1 - 1} \prod_{k=2}^{K} \theta_k^{\tilde{\alpha}_k - 1} \, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1 + 1) \prod_{k=2}^{K} \Gamma(\tilde{\alpha}_k)}{\Gamma(\tilde{\alpha}_1 + 1 + \sum_{k=2}^{K} \tilde{\alpha}_k)} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\Gamma(\tilde{\alpha}_1 + 1 + \sum_{k=2}^{K} \tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1 + 1) \prod_{k=2}^{K} \Gamma(\tilde{\alpha}_k)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\Gamma(1 + \sum_{k=1}^{K} \tilde{\alpha}_k)} \times \frac{\Gamma(\tilde{\alpha}_1 + 1)}{\Gamma(\tilde{\alpha}_1)} = \frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{(\sum_{k=1}^{K} \tilde{\alpha}_k)\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)} \times \frac{\tilde{\alpha}_1 \Gamma(\tilde{\alpha}_1)}{\Gamma(\tilde{\alpha}_1)} \\
&= \frac{\tilde{\alpha}_1}{\sum_{k=1}^{K} \tilde{\alpha}_k} = \frac{M_1 + \alpha_1}{\sum_{k=1}^{K} (M_k + \alpha_k)}
\end{aligned}$$

Discrete Bayesian network



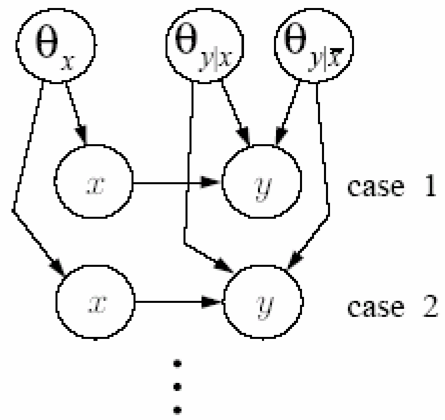
$$P\Big(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big) = \prod_{k=1}^{K} \theta_{n,\pi_n,k} \mathcal{I}(\mathcal{D}_{n,t} = k)$$

Discrete Bayesian Networks



$$P\Big(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big) = \prod_{k=1}^{K} \theta_{n,\pi_n,k} \mathcal{I}(\mathcal{D}_{n,t} = k)$$

Discrete Bayesian Networks



$$P\Big(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big) = \prod_{k=1}^{K} \theta_{n,\pi_n,k} \mathcal{I}(\mathcal{D}_{n,t} = k)$$

$$\begin{aligned} P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta}) &= \prod_{n=1}^{N}\prod_{t=1}^{M} P\Big(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t) = \mathcal{D}_{(\pi_n,t)}, \boldsymbol{\theta}_{n,\pi_n}\Big) \\ &= \prod_{n=1}^{N}\prod_{t=1}^{M}\prod_{k=1}^{K} \theta_{n,\pi_n,k} \mathcal{I}(\mathcal{D}_{n,t} = k) \\ &= \prod_{n}\prod_{\pi_n}\prod_{k} [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}} \end{aligned}$$

Parameter modularity:

$$P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}) \;=\; P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G}') \;:=\; P(\boldsymbol{\theta}_{n,\pi_n})$$

Parameter independence:

$$P(\boldsymbol{\theta}|\mathcal{G}) \;=\; \prod_{n=1}^{N} P(\boldsymbol{\theta}_{n,\pi_n}|\mathcal{G})$$

Conjugacy:

$$P(\boldsymbol{\theta}_{n,\pi_n}) \;=\; Dir(\boldsymbol{\theta}_{n,\pi_n}|\boldsymbol{\alpha}_{n,\pi_n}) \;=\; \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\prod_k \Gamma(\alpha_{n,\pi_n,k})} \prod_k [\theta_{n,\pi_n,k}]^{\alpha_{n,\pi_n,k}-1}$$

Likelihood:

$$P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta}) = \prod_n \prod_{\pi_n} \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}}$$

Recall:

$$Dir(\boldsymbol{\theta}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Marginal likelihood:

$$\begin{aligned} P(\mathcal{D}|\mathcal{G}) &= \int P(\mathcal{D}|\boldsymbol{\theta},\mathcal{G}) P(\boldsymbol{\theta}|\mathcal{G}) d\boldsymbol{\theta} \\ &= \prod_n \prod_{\pi_n} \int \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}} Dir(\boldsymbol{\theta}_{n,\pi_n}|\boldsymbol{\alpha}_{n,\pi_n}) d\boldsymbol{\theta}_{n,\pi_n} \\ &= \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\prod_k \Gamma(\alpha_{n,\pi_n,k})} \int \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k} + \alpha_{n,\pi_n,k} - 1} d\boldsymbol{\theta}_{n,\pi_n} \end{aligned}$$

Likelihood:

$$P(\mathcal{D}|\mathcal{G},\boldsymbol{\theta}) = \prod_n \prod_{\pi_n} \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}}$$

Recall:

$$Dir(\boldsymbol{\theta}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

$$\int_{\boldsymbol{\theta}\in\mathcal{S}} \prod_{k=1}^{K} \theta_k^{\alpha_k-1} d\boldsymbol{\theta} = \left(\frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)}\right)$$

Marginal likelihood:

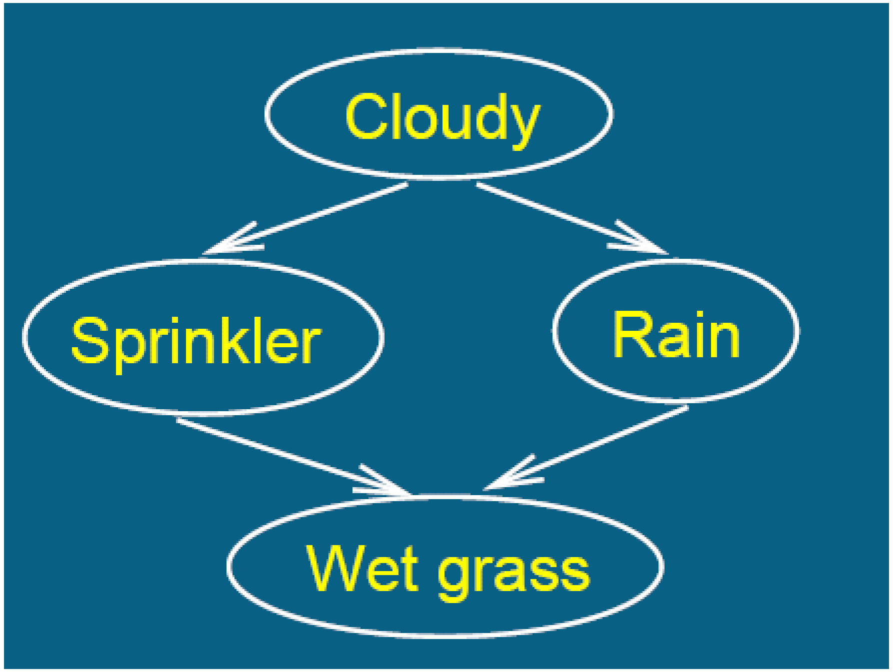
$$\begin{aligned}
P(\mathcal{D}|\mathcal{G}) &= \int P(\mathcal{D}|\boldsymbol{\theta},\mathcal{G}) P(\boldsymbol{\theta}|\mathcal{G}) d\boldsymbol{\theta} \\
&= \prod_n \prod_{\pi_n} \int \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}} Dir(\boldsymbol{\theta}_{n,\pi_n}|\boldsymbol{\alpha}_{n,\pi_n}) d\boldsymbol{\theta}_{n,\pi_n} \\
&= \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\prod_k \Gamma(\alpha_{n,\pi_n,k})} \int \prod_k [\theta_{n,\pi_n,k}]^{M_{n,\pi_n,k}+\alpha_{n,\pi_n,k}-1} d\boldsymbol{\theta}_{n,\pi_n} \\
&= \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\prod_k \Gamma(\alpha_{n,\pi_n,k})} \times \frac{\prod_k \Gamma(M_{n,\pi_n,k}+\alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k}+\alpha_{n,\pi_n,k}])} \\
&= \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k}+\alpha_{n,\pi_n,k}])} \prod_k \frac{\Gamma(M_{n,\pi_n,k}+\alpha_{n,\pi_n,k})}{\Gamma(\alpha_{n,\pi_n,k})}
\end{aligned}$$

BDe score

$$P(\mathcal{D}|\mathcal{G}) = \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k} + \alpha_{n,\pi_n,k}])} \prod_k \frac{\Gamma(M_{n,\pi_n,k} + \alpha_{n,\pi_n,k})}{\Gamma(\alpha_{n,\pi_n,k})}$$

BDe score

$$P(\mathcal{D}|\mathcal{G}) = \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k} + \alpha_{n,\pi_n,k}])} \prod_k \frac{\Gamma(M_{n,\pi_n,k} + \alpha_{n,\pi_n,k})}{\Gamma(\alpha_{n,\pi_n,k})}$$



n	π_n	k
Clouds	–	on , off
Sprinkler	Clouds = off	on , off
Sprinkler	Clouds = on	on , off
Rain	Clouds = off	on , off
Rain	Clouds = on	on , off
Wet grass	Clouds= off, sprinkler= off	on, off
Wet grass	Clouds= off, sprinkler= on	on, off
Wet grass	Clouds= on, sprinkler= off	on, off
Wet grass	Clouds= on, sprinkler= on	on, off

BDe score

$$P(\mathcal{D}|\mathcal{G}) = \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k} + \alpha_{n,\pi_n,k}])} \prod_k \frac{\Gamma(M_{n,\pi_n,k} + \alpha_{n,\pi_n,k})}{\Gamma(\alpha_{n,\pi_n,k})}$$

How to set the hyperparameters?

BDe score

$$P(\mathcal{D}|\mathcal{G}) = \prod_n \prod_{\pi_n} \frac{\Gamma(\sum_k \alpha_{n,\pi_n,k})}{\Gamma(\sum_k [M_{n,\pi_n,k} + \alpha_{n,\pi_n,k}])} \prod_k \frac{\Gamma(M_{n,\pi_n,k} + \alpha_{n,\pi_n,k})}{\Gamma(\alpha_{n,\pi_n,k})}$$

How to set the hyperparameters?

Prior network $\mathcal{G}_0$ and equivalent sampe size M_0. To satisfy event equivalence:

$$\alpha_{n,\pi_n,k} = M_0 P(X_n = k, \pi_n | \mathcal{G}_0)$$

Uninformative case:

$$\alpha_{n,\pi_n,k} = \frac{M_0}{K_n \times \sharp[\pi_n]}$$

Number of discretization levels for node n

Number of parent configurations for node n

BDe: UAI 1994

Learning Bayesian Networks: The Combination of Knowledge and Statistical Data

David Heckerman **Dan Geiger*** **David M. Chickering**
Microsoft Research, Bldg 9S
Redmond, WA 98052-6399
heckerma@microsoft.com, dang@cs.technion.ac.il, dmax@cs.ucla.edu

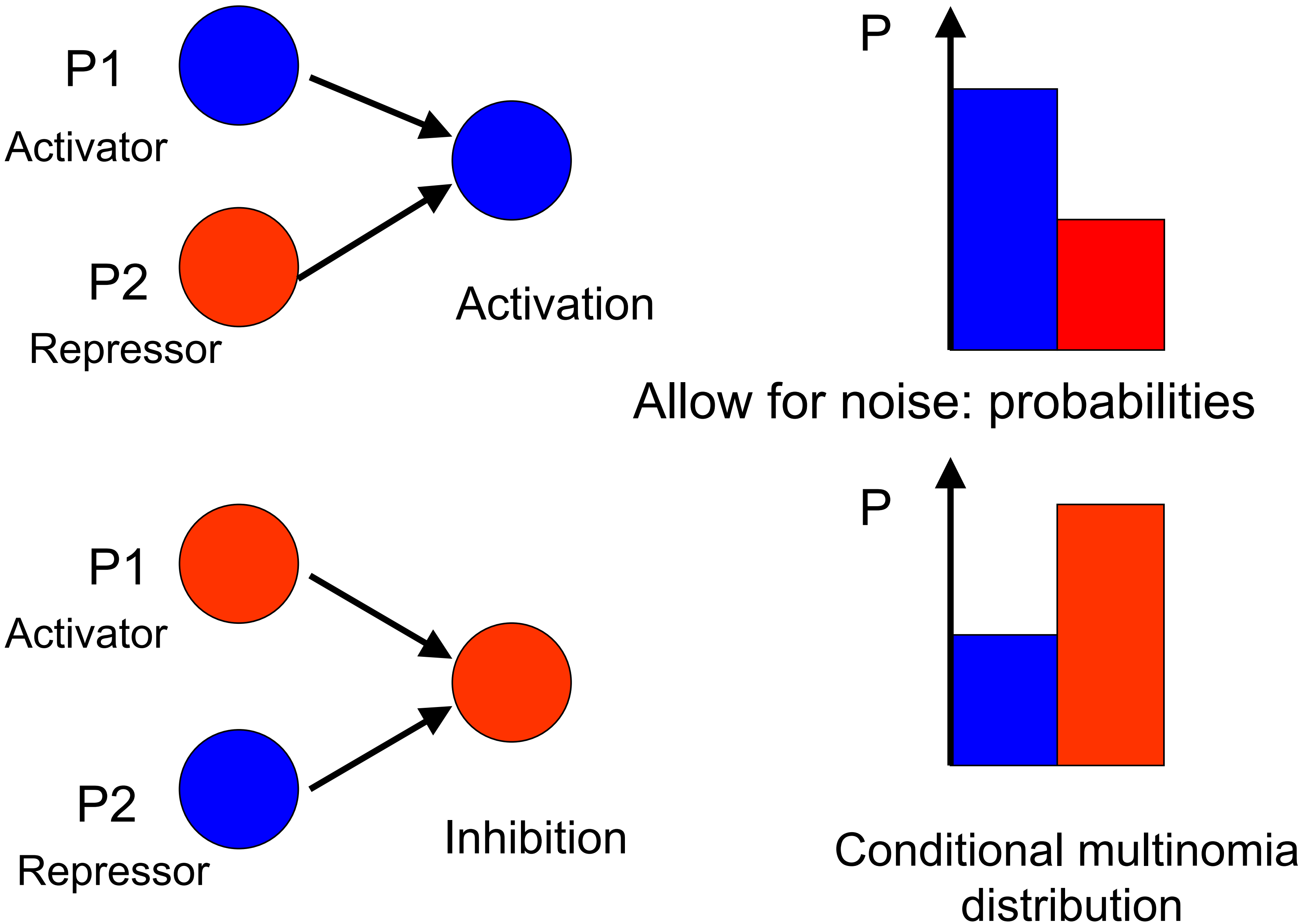
BGe: UAI 1995

Learning Bayesian Networks: A Unification for Discrete and Gaussian Domains

David Heckerman and Dan Geiger*

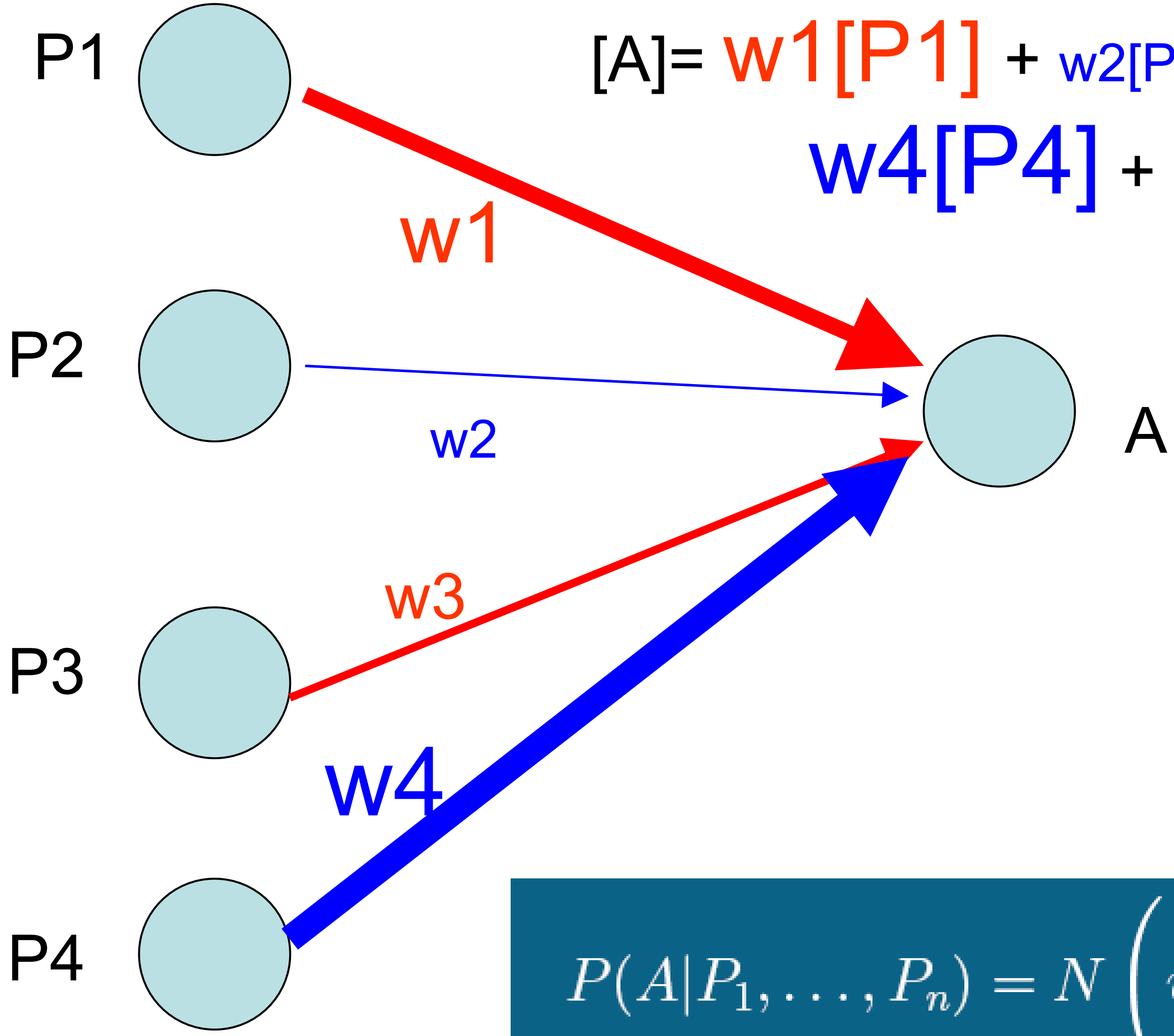
Microsoft Research, Bldg 9S/1
Redmond 98052-6399, WA
heckerma@microsoft.com,dang@cs.technion.ac.il

BDe: Nonlinear discretized model



BGe: Linear model

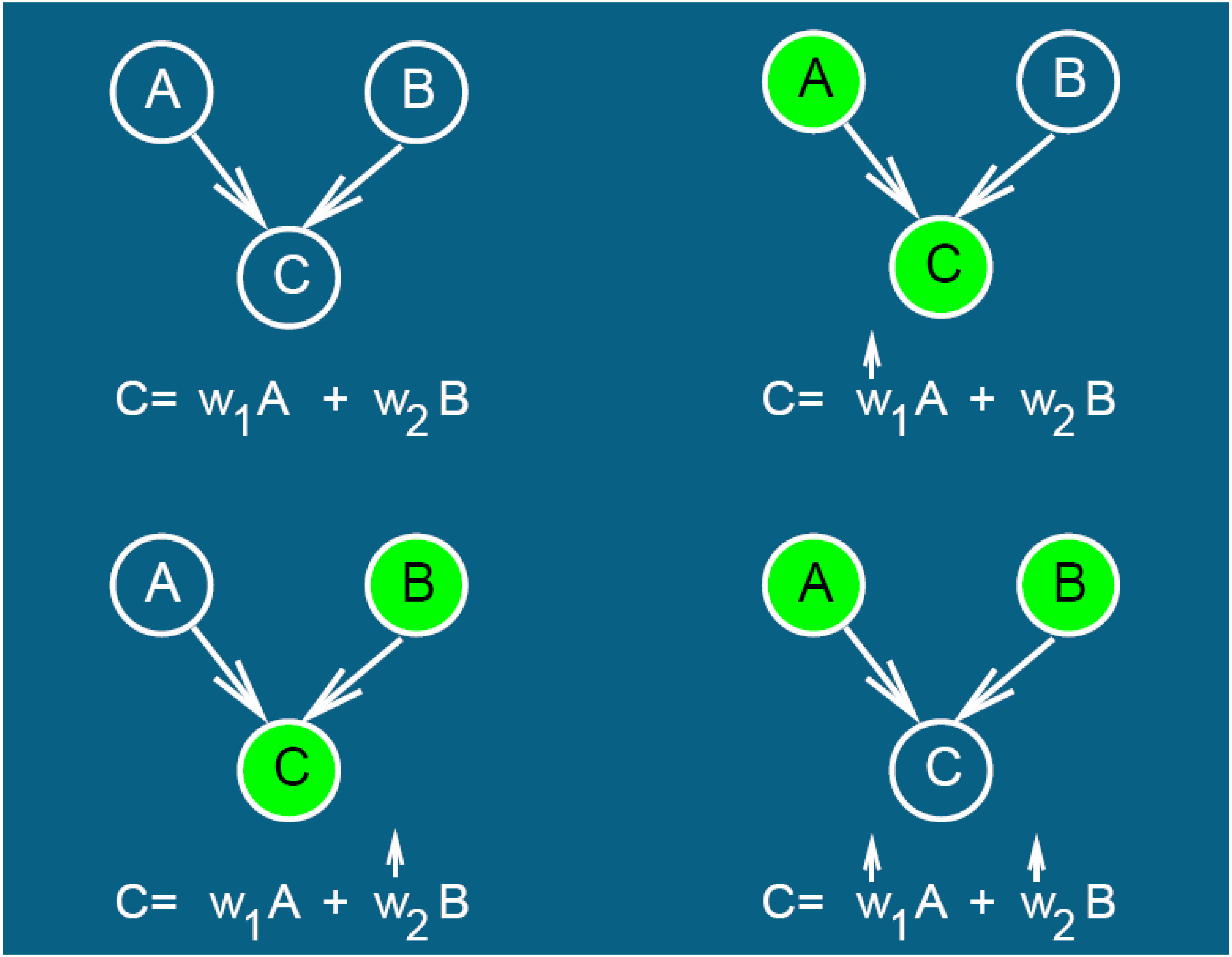
[A]= w1[P1] + w2[P2] + w3[P3] + w4[P4] + noise



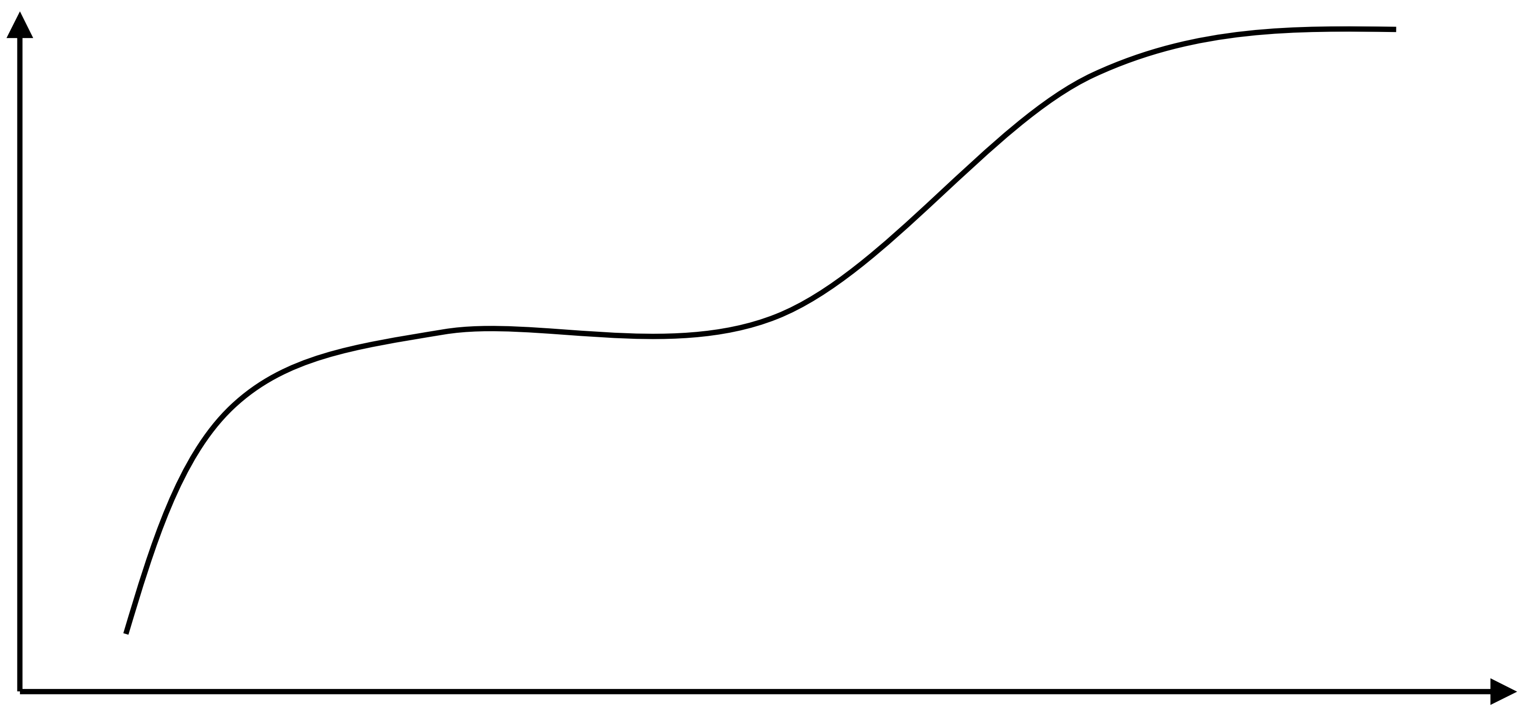
$$P(A|P_1, \ldots, P_n) = N\left(w_0 + \sum_{i=1}^{n} w_i P_i, \sigma^2\right)$$

Pros and cons of the two models

Linear Gaussian model	Multinomial model
• Restriction to linear processes	• Nonlinear model
• Original data → no information loss	• Discretization → information loss

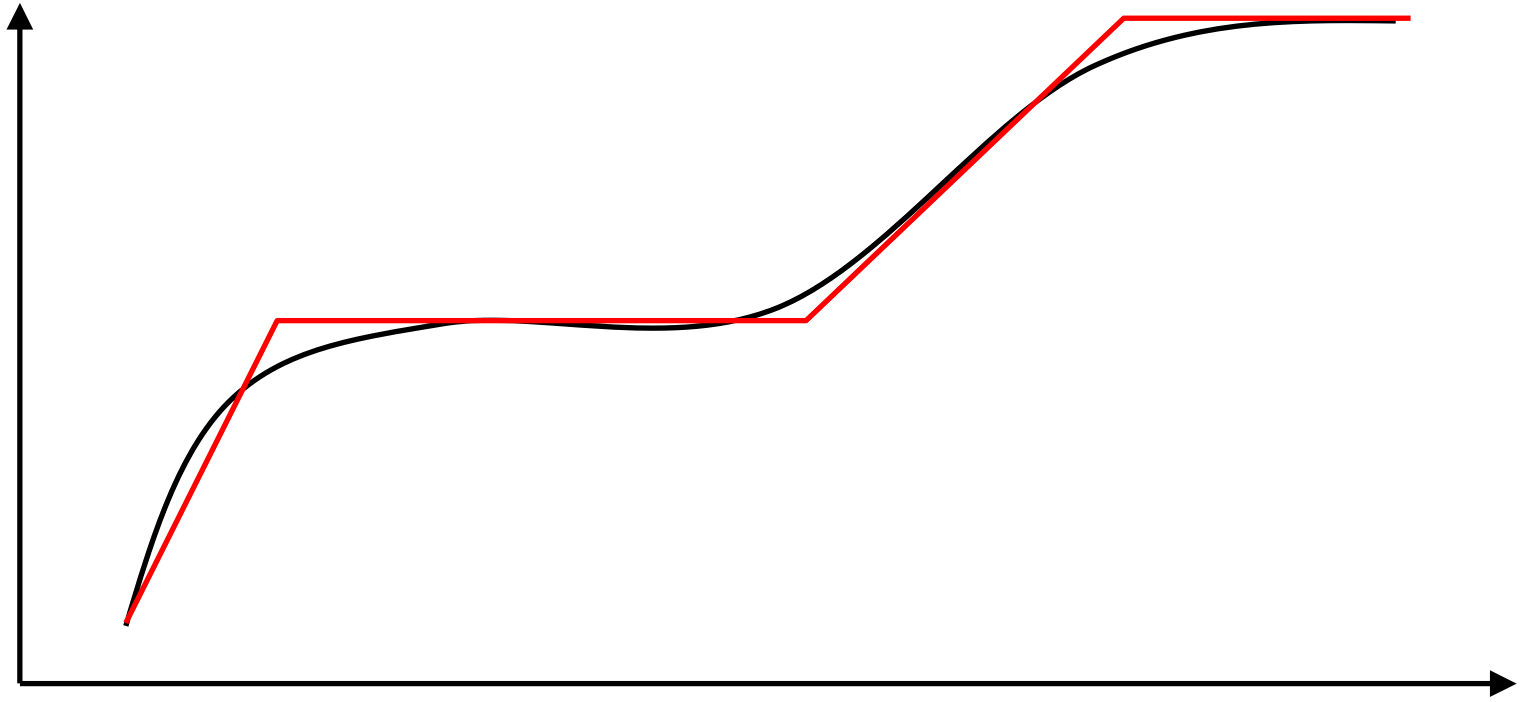

A
B
C
C= w1A + w2B
A
B
C
C= w1A + w2B
A
B
C
C= w1A + w2B
A
B
C
C= w1A + w2B

Can we get an approximate nonlinear model without data discretization?



Can we get an approximate nonlinear model without data discretization?

Idea: piecewise linear model



Non-stationary continuous dynamic Bayesian networks

Marco Grzegorczyk
Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
grzegorczyk@statistik.tu-dortmund.de

Dirk Husmeier
Biomathematics & Statistics Scotland (BioSS)
JCMB, The King's Buildings, Edinburgh EH93JZ, United Kingdom
dirk@bioss.ac.uk

NIPS 2009