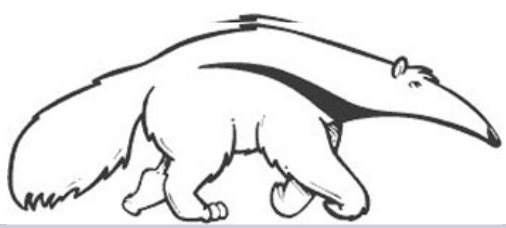
Learning Undirected Models

Learning in Graphical Models

Prof. Alexander Ihler



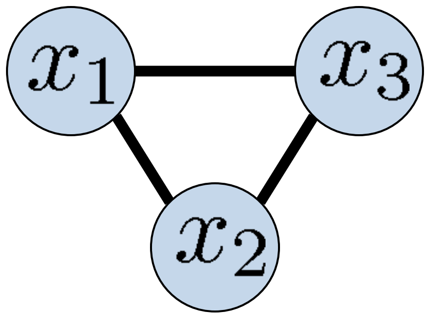
BREN:ICS
INFORMATION AND COMPUTER SCIENCES

UNIVERSITY of CALIFORNIA IRVINE

Learning non-trees

- No closed form solution

$$\mathcal{L} = \sum_j \big[\sum_\alpha \log f_\alpha(x_\alpha^j) - \log Z(f) \big]$$



$$p(x) = \frac{1}{Z} f_{12}(x_1, x_2) f_{13}(x_1, x_3) f_{23}(x_2, x_3)$$

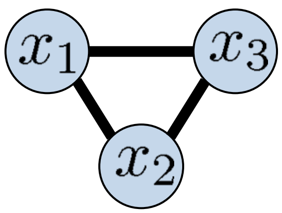
- Use exponential family ideas:
  - Exp family: $\log f_\alpha(x) = \sum_k \theta_{\alpha;k} u_{\alpha;k}(x)$
  - Gradient $$\frac{\partial \mathcal{L}}{\partial \theta_{\alpha;k}} = \sum_j u_{\alpha;k}(x^j) - m\mathbb{E}[u_{\alpha;k}(x)]$$
- Then:
  - Gradient ascent
  - Coordinate updates

Both require **inference**:
need expectation under the current model

Learning & inference

- A conversion between representations

Model



$$p(x) = \frac{1}{Z} f_{12}(x_1, x_2) f_{13}(x_1, x_3) f_{23}(x_2, x_3)$$

Compact: here, 3 d^2 parameters
Prior knowledge (structure)
Restricted set of distributions

Requires inference to compute probabilities, e.g., p(x1)

Empirical distribution

$$x^{(1)} = [0 \;\; 1 \;\; 1]$$
$$x^{(2)} = [0 \;\; 0 \;\; 0]$$
$$x^{(3)} = [1 \;\; 0 \;\; 1]$$
$$\vdots$$

Large (min of m, d^n)
No structure, no assumptions

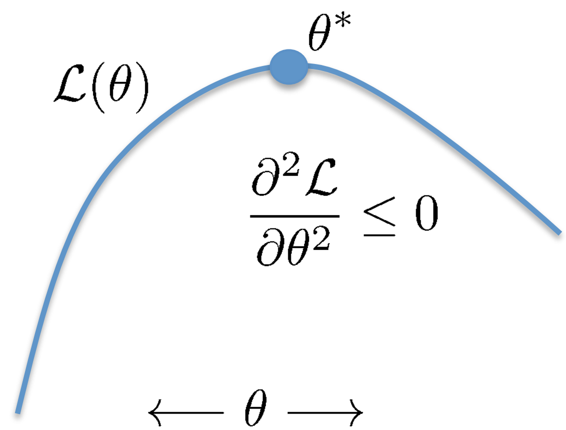
Trivial to compute probabilities

ML for exponential families

- The log-likelihood is concave:

$$\mathcal{L} = \sum_j \Big[\sum_\alpha \theta_\alpha u_\alpha(x) - \log Z(\theta) \Big]$$

$$\frac{\partial \mathcal{L}}{\partial \theta_a} = \sum_j u_a(x^{(j)}) - \mathbb{E}[u_a(x)]$$



$$\frac{\partial^2 \mathcal{L}}{\partial \theta_a \partial \theta_b} = -\frac{\partial}{\partial \theta_b} \sum_x \exp[\sum_\alpha \theta_\alpha u_\alpha(x) - A(\theta)] u_a(x)$$

$$= -\sum_x \exp[\sum_\alpha \theta_\alpha u_\alpha(x) - A(\theta)](u_b(x) - \mathbb{E}[u_b(x)])\, u_a(x)$$

$$= -\mathbb{E}[u_b(x) u_a(x)] + \mathbb{E}[u_b(x)]\, \mathbb{E}[u_a(x)] \qquad = \mathrm{Cov}(u_a, u_b)$$

$\nabla^2 \mathcal{L}$ is negative semi-definite: concave function

No local maxima – if overcomplete, may be multiple global maxima

Iterative Scaling

Model $p(x\,;\,\theta) = \exp\big[\sum_i \theta_i u_i(x) - A(\theta)\big]$

Likelihood $\frac{1}{m}\mathcal{L} = \sum_i \theta_i \hat{\mu}_i - A(\theta)$

Empirical moments

$$\hat{\mu}_i = \frac{1}{m}\sum_j u_i(x^{(j)})$$

Choose a subset of features S:

$$\forall x, \quad u_i(x) \geq 0 \ \text{ and } \ \sum_{i \in S} u_i(x) = 1$$

and consider an update

$$\theta \to \theta' = \theta + \Delta\theta$$

Then, $\frac{1}{m}\Delta\mathcal{L} = \sum_i \Delta\theta_i \hat{\mu}_i - \Delta A(\theta) \quad \geq \sum_i \Delta\theta_i \hat{\mu}_i - \sum_i \exp[\Delta\theta_i]\mu_i + 1$

Lower-bound the likelihood:

$$\Delta A(\theta) = \log \frac{Z(\theta')}{Z(\theta)} \leq \frac{Z(\theta')}{Z(\theta)} - 1$$

$$\begin{aligned}
\frac{Z(\theta')}{Z(\theta)} &= \sum_x p(x;\theta) \exp\big[\sum_i \Delta\theta_i\, u_i(x)\big] \\
&\leq \sum_x p(x;\theta) \sum_i u_i(x) e^{\Delta\theta_i} \qquad \text{(Jensen's)} \\
&= \sum_i e^{\Delta\theta_i} \mu_i
\end{aligned}$$

Iterative Scaling

Model $p(x\,;\,\theta) = \exp\big[\sum_i \theta_i u_i(x) - A(\theta)\big]$

Likelihood $\frac{1}{m}\mathcal{L} = \sum_i \theta_i \hat{\mu}_i - A(\theta)$

Empirical moments

$$\hat{\mu}_i = \frac{1}{m}\sum_j u_i(x^{(j)})$$

Choose a subset of features S:

$$\forall x, \quad u_i(x) \geq 0 \ \text{ and } \ \sum_{i \in S} u_i(x) = 1$$

and consider an update

$$\theta \to \theta' = \theta + \Delta\theta$$

Then, $\frac{1}{m}\Delta\mathcal{L} = \sum_i \Delta\theta_i \hat{\mu}_i - \Delta A(\theta) \quad \geq \sum_i \Delta\theta_i \hat{\mu}_i - \sum_i \exp[\Delta\theta_i]\mu_i + 1$

Optimize the lower bound:

$$\frac{\partial}{\partial a}(a\hat{\mu} - e^a\mu) = \hat{\mu} - e^a\mu = 0$$

$$\Rightarrow a = \log(\hat{\mu}/\mu)$$

Gives update: $\Delta\theta^* = \log(\hat{\mu}_i/\mu_i)$

To optimize, iterate:

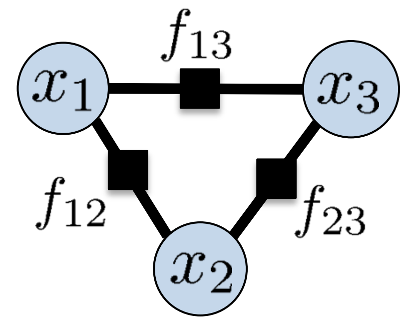
(1) choose set S and compute $\mu_i = \mathbb{E}_p\big[u_i(x)\big]$

(2) update parameters $\theta_i \longleftarrow \theta_i + \log \frac{\hat{\mu}_i}{\mu_i}$

or equivalently:

$$p(x\,;\,\theta) \longleftarrow \frac{1}{Z} p(x\,;\,\theta) \prod_i \Big(\frac{\hat{\mu}_i}{\mu_i}\Big)^{u_i(x)}$$

Iterative Proportional Fitting



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

Model $$p^{(t)}(x) = \frac{1}{Z} \prod_{\alpha} f_{\alpha}^{(t)}(x_\alpha)$$

Empirical moments $$\hat{p}(x_\alpha) = \frac{1}{m} \sum_j \delta(x_\alpha = x_\alpha^j)$$

The features $S_\alpha = \{ u_{\alpha,k}(x) = \delta(x_\alpha = k) \}$ have exactly one $u_{\alpha,k}(x) = 1$ for all values x

Iterate to convergence:

for each clique α :

(1) compute model's marginal

$$p^{(t)}(x_\alpha) = \sum_{x \setminus x_\alpha} p^{(t)}(x)$$

(2) update model parameters

$$f_\alpha^{(t+1)}(x_\alpha) = f_\alpha^{(t)}(x_\alpha) \cdot \frac{\hat{p}_\alpha(x_\alpha)}{p^{(t)}(x_\alpha)}$$

Notes:

(1) Moment matching

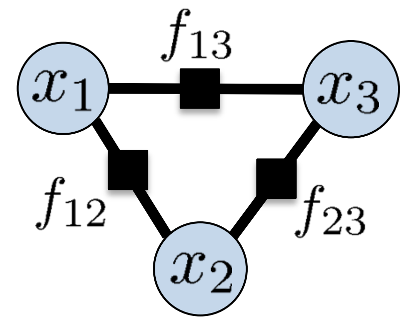
=> sets derivative of L w.r.t. $f_\alpha(x)$ to zero

=> coordinate ascent on L

(2) If L is strictly convex

=> procedure converges to the MLE

Iterative Proportional Fitting



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

```
f12 = factor( [1 2], ones(3,3);   % initialize model
f13 = factor( [1 3], ones(3,3);
f23 = factor( [2 3], ones(3,3);

ph12 = empirical(f12, D(:,[1 2]);   % compute empirical
ph13 = empirical(f13, D(:,[1 3]);   %    moments
ph23 = empirical(f23, D(:,[2 3]);
```

Model

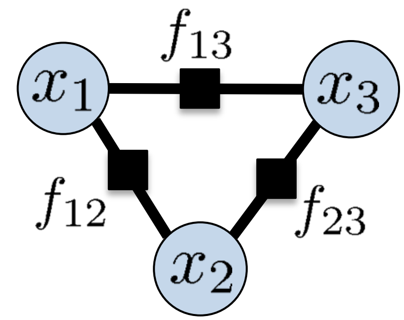
$$f_{12} \quad \begin{array}{cc} & - x_2 - \\ | \\ x_1 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ | \end{array} \qquad f_{13} \quad \begin{array}{cc} & - x_3 - \\ | \\ x_1 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ | \end{array} \qquad f_{23} \quad \begin{array}{cc} & - x_3 - \\ | \\ x_2 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ | \end{array}$$

Empirical

$$\hat{p}_{12} \begin{pmatrix} 0.249 & 0.002 & 0.311 \\ 0.017 & 0.024 & 0.015 \\ 0.029 & 0.348 & 0.000 \end{pmatrix} \qquad \hat{p}_{13} \begin{pmatrix} 0.136 & 0.118 & 0.309 \\ 0.003 & 0.029 & 0.025 \\ 0.348 & 0.014 & 0.015 \end{pmatrix} \qquad \hat{p}_{23} \begin{pmatrix} 0.008 & 0.014 & 0.274 \\ 0.351 & 0.024 & 0.000 \\ 0.129 & 0.123 & 0.074 \end{pmatrix}$$

Iterative Proportional Fitting

```
f12 = factor( [1 2], ones(3,3);      ... % initialize model
ph12 = empirical(f12, D(:,[1 2]); ... % compute empirical

p12 = f12 * sum(f13*f23, 3);         % compute marginal
p12 = normalize(p12);
```



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

Model

$$\begin{array}{cc} f_{12} & -x_2- \\ |\,x_1\,| & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \quad \begin{array}{cc} f_{13} & -x_3- \\ |\,x_1\,| & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \quad \begin{array}{cc} f_{23} & -x_3- \\ |\,x_2\,| & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

$$\begin{pmatrix} 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \end{pmatrix}$$

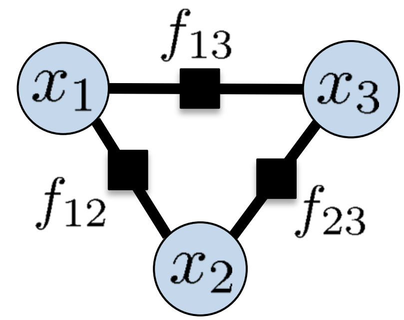
$$p_{12} \propto f_{12} \sum_{x_3} f_{13}\, f_{23}$$

Empirical

$$\hat{p}_{12} \begin{pmatrix} 0.249 & 0.002 & 0.311 \\ 0.017 & 0.024 & 0.015 \\ 0.029 & 0.348 & 0.000 \end{pmatrix} \quad \hat{p}_{13} \begin{pmatrix} 0.136 & 0.118 & 0.309 \\ 0.003 & 0.029 & 0.025 \\ 0.348 & 0.014 & 0.015 \end{pmatrix} \quad \hat{p}_{23} \begin{pmatrix} 0.008 & 0.014 & 0.274 \\ 0.351 & 0.024 & 0.000 \\ 0.129 & 0.123 & 0.074 \end{pmatrix}$$

Iterative Proportional Fitting

```
f12 = factor( [1 2], ones(3,3);     ... % initialize model
ph12 = empirical(f12, D(:,[1 2]); ... % compute empirical

p12 = f12 * sum(f13*f23, 3);          % compute marginal
p12 = normalize(p12);

f12 = f12 * ph12 / p12;               % IPF update
```



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

Model

$$f_{12}\ \begin{array}{c|c} & -x_2- \\ \hline x_1 & \begin{pmatrix} 2.245 & 0.025 & 2.803 \\ 0.160 & 0.218 & 0.141 \\ 0.261 & 3.140 & 0.003 \end{pmatrix} \end{array} \quad f_{13}\ \begin{array}{c|c} & -x_3- \\ \hline x_1 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \quad f_{23}\ \begin{array}{c|c} & -x_3- \\ \hline x_2 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

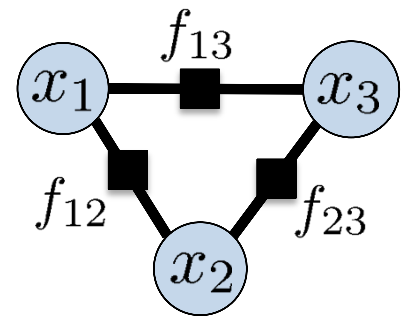
$$\times \frac{\hat{p}_{12}}{p_{12}} \qquad \begin{pmatrix} 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \end{pmatrix} \qquad p_{12} \propto f_{12} \sum_{x_3} f_{13}\, f_{23}$$

Empirical

$$\hat{p}_{12} \begin{pmatrix} 0.249 & 0.002 & 0.311 \\ 0.017 & 0.024 & 0.015 \\ 0.029 & 0.348 & 0.000 \end{pmatrix} \quad \hat{p}_{13} \begin{pmatrix} 0.136 & 0.118 & 0.309 \\ 0.003 & 0.029 & 0.025 \\ 0.348 & 0.014 & 0.015 \end{pmatrix} \quad \hat{p}_{23} \begin{pmatrix} 0.008 & 0.014 & 0.274 \\ 0.351 & 0.024 & 0.000 \\ 0.129 & 0.123 & 0.074 \end{pmatrix}$$

Iterative Proportional Fitting

```
f12 = factor( [1 2], ones(3,3);      … % initialize model
ph12 = empirical(f12, D(:,[1 2]); … % compute empirical

p13 = f13 * sum(f12*f23, 2);          % compute marginal
p13 = normalize(p13);
```



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

Model

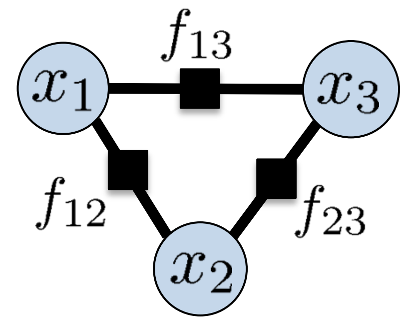
$$f_{12} \quad \begin{array}{c|c} & - x_2 - \\ \hline x_1 & \begin{pmatrix} 2.245 & 0.025 & 2.803 \\ 0.160 & 0.218 & 0.141 \\ 0.261 & 3.140 & 0.003 \end{pmatrix} \end{array} \qquad f_{13} \quad \begin{array}{c|c} & - x_3 - \\ \hline x_1 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \qquad f_{23} \quad \begin{array}{c|c} & - x_3 - \\ \hline x_2 & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

$$\begin{pmatrix} 0.187 & 0.187 & 0.187 \\ 0.019 & 0.019 & 0.019 \\ 0.126 & 0.126 & 0.126 \end{pmatrix} \qquad p_{13} \propto f_{13} \sum_{x_2} f_{12}\, f_{23}$$

Empirical

$$\hat{p}_{12} \begin{pmatrix} 0.249 & 0.002 & 0.311 \\ 0.017 & 0.024 & 0.015 \\ 0.029 & 0.348 & 0.000 \end{pmatrix} \qquad \hat{p}_{13} \begin{pmatrix} 0.136 & 0.118 & 0.309 \\ 0.003 & 0.029 & 0.025 \\ 0.348 & 0.014 & 0.015 \end{pmatrix} \qquad \hat{p}_{23} \begin{pmatrix} 0.008 & 0.014 & 0.274 \\ 0.351 & 0.024 & 0.000 \\ 0.129 & 0.123 & 0.074 \end{pmatrix}$$

Iterative Proportional Fitting



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

```
f12 = factor( [1 2], ones(3,3);      ... % initialize model
ph12 = empirical(f12, D(:,[1 2]); ... % compute empirical

p13 = f13 * sum(f12*f23, 2);          % compute marginal
p13 = normalize(p13);

f13 = f13 * ph13 / p13;               % IPF update
```

Model

$$f_{12} \quad \begin{array}{c} \\ x_1 \end{array} \begin{array}{c} - x_2 - \\ \begin{pmatrix} 2.245 & 0.025 & 2.803 \\ 0.160 & 0.218 & 0.141 \\ 0.261 & 3.140 & 0.003 \end{pmatrix} \end{array} \qquad f_{13} \quad \begin{array}{c} \\ x_1 \end{array} \begin{array}{c} - x_3 - \\ \begin{pmatrix} 0.724 & 0.631 & 1.644 \\ 0.189 & 1.504 & 1.306 \\ 2.765 & 0.113 & 0.121 \end{pmatrix} \end{array} \qquad f_{23} \quad \begin{array}{c} \\ x_2 \end{array} \begin{array}{c} - x_3 - \\ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

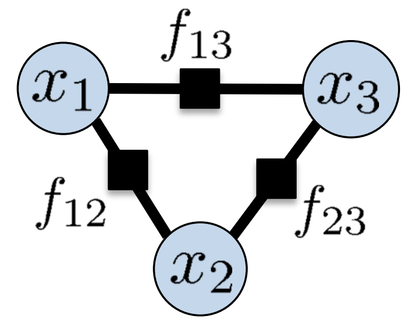
$$\times \frac{\hat{p}_{13}}{p_{13}} \qquad \begin{pmatrix} 0.187 & 0.187 & 0.187 \\ 0.019 & 0.019 & 0.019 \\ 0.126 & 0.126 & 0.126 \end{pmatrix} \qquad p_{13} \propto f_{13} \sum_{x_2} f_{12} \, f_{23}$$

Empirical

$$\hat{p}_{12} \begin{pmatrix} 0.249 & 0.002 & 0.311 \\ 0.017 & 0.024 & 0.015 \\ 0.029 & 0.348 & 0.000 \end{pmatrix} \qquad \hat{p}_{13} \begin{pmatrix} 0.136 & 0.118 & 0.309 \\ 0.003 & 0.029 & 0.025 \\ 0.348 & 0.014 & 0.015 \end{pmatrix} \qquad \hat{p}_{23} \begin{pmatrix} 0.008 & 0.014 & 0.274 \\ 0.351 & 0.024 & 0.000 \\ 0.129 & 0.123 & 0.074 \end{pmatrix}$$

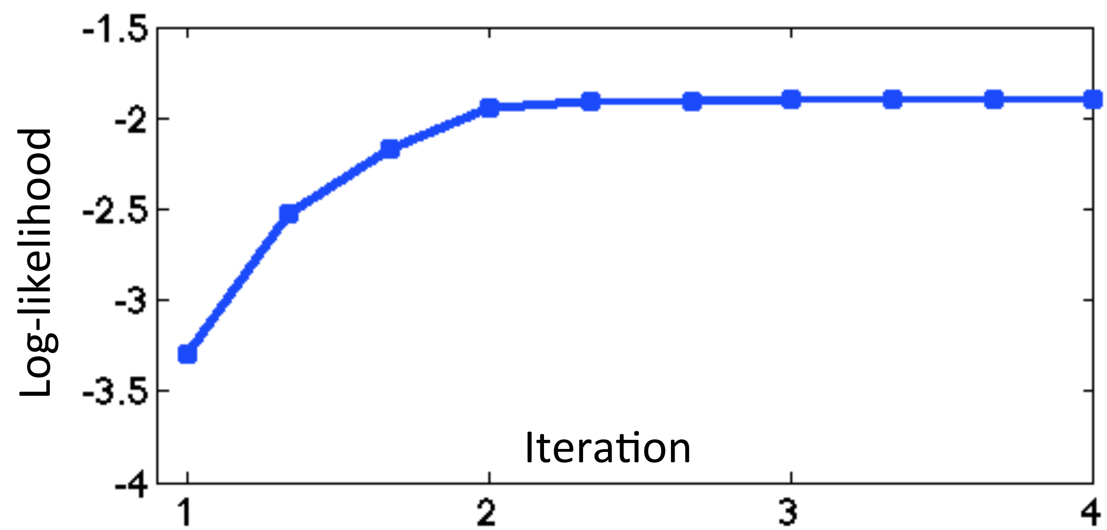
Iterative Proportional Fitting

```
f12 = factor( [1 2], ones(3,3);      ... % initialize model
ph12 = empirical(f12, D(:,[1 2]); ... % compute empirical

p13 = f13 * sum(f12*f23, 2);          % compute marginal
p13 = normalize(p13);

f13 = f13 * ph13 / p13;               % IPF update
```



$$p(x) \propto f_{12} \cdot f_{13} \cdot f_{23}$$

- Coordinate ascent on the log-likelihood
  - Each step (re-)fits a set of moments



IPF in Gaussian distributions

- Recall: structure expressed as sparse inverse covariance

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Let's assume mean zero (easy to enforce), J = inv(Sigma)
- IPF: find $p(x_C)$ for some C; set $p(x_C) = \text{emp}(x_C) / p(x_C)$

$$p(x_a, x_b) = \mathcal{N}([x_a, x_b]\,;\,\Sigma_{ab})$$

Invert J & find [a,b] sub-matrix

$$\hat{p}(x_a, x_b) = \mathcal{N}([x_a, x_b]\,;\,\hat{\Sigma}_{ab})$$

Want: empirical covariance on a,b

IPF update:

$$p(x) \leftarrow p(x)\frac{\hat{p}(x_a, x_b)}{p(x_a, x_b)} \quad \Rightarrow \quad J_{ab} \leftarrow J_{ab} + (\hat{\Sigma}_{ab})^{-1} - (\Sigma_{ab})^{-1}$$

Gaussian IPF

- Initialize; choose a factor $f(x_i,x_j)$
- Update to match empirical marginal

Gaussian distribution:

$$\mathcal{N}(x\,;\,\mu,\Sigma) = \mathcal{N}^{-1}(x\,;\,h,J)$$

where $J = \Sigma^{-1}$

$h = \mu\,\Sigma^{-1}$

$$J = \begin{pmatrix} 1.000 & -0.250 & \mathbf{0} & -0.250 \\ -0.250 & 1.000 & -0.250 & \mathbf{0} \\ \mathbf{0} & -0.250 & 1.000 & -0.250 \\ -0.250 & \mathbf{0} & -0.250 & 1.000 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 1.167 & 0.333 & 0.167 & 0.333 \\ 0.333 & 1.167 & 0.333 & 0.167 \\ 0.167 & 0.333 & 1.167 & 0.333 \\ 0.333 & 0.167 & 0.333 & 1.167 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 0.912 & -0.239 & 0.295 & 0.002 \\ -0.239 & 1.041 & -0.322 & 0.261 \\ 0.295 & -0.322 & 0.928 & -0.053 \\ 0.002 & 0.261 & -0.053 & 0.873 \end{pmatrix}$$

$$\Sigma_{[12]} = \begin{pmatrix} 1.167 & 0.333 \\ 0.333 & 1.167 \end{pmatrix} \qquad \hat{\Sigma}_{[12]} = \begin{pmatrix} 0.912 & -0.239 \\ -0.239 & 1.041 \end{pmatrix}$$

$$(\Sigma_{[12]})^{-1} = \begin{pmatrix} 0.933 & -0.267 \\ -0.267 & 0.933 \end{pmatrix} \qquad (\hat{\Sigma}_{[12]})^{-1} = \begin{pmatrix} 1.167 & 0.268 \\ 0.268 & 1.021 \end{pmatrix}$$

$$J = \begin{pmatrix} 1.233 & 0.285 & \mathbf{0} & -0.250 \\ 0.285 & 1.088 & -0.250 & \mathbf{0} \\ \mathbf{0} & -0.250 & 1.000 & -0.250 \\ -0.250 & \mathbf{0} & -0.250 & 1.000 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.912 & -0.239 & -0.003 & 0.227 \\ -0.239 & 1.041 & 0.261 & 0.005 \\ -0.003 & 0.261 & 1.136 & 0.283 \\ 0.227 & 0.005 & 0.283 & 1.127 \end{pmatrix}$$

Gaussian IPF

- Initialize; choose a factor $f(x_i,x_j)$
- Update to match empirical marginal

Gaussian distribution:

$$\mathcal{N}(x\,;\,\mu,\Sigma) = \mathcal{N}^{-1}(x\,;\,h,J)$$

where $J = \Sigma^{-1}$

$h = \mu\,\Sigma^{-1}$

$$J = \begin{pmatrix} 1.233 & 0.285 & \mathbf{0} & -0.250 \\ 0.285 & 1.088 & -0.250 & \mathbf{0} \\ \mathbf{0} & -0.250 & 1.000 & -0.250 \\ -0.250 & \mathbf{0} & -0.250 & 1.000 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 0.912 & -0.239 & -0.003 & 0.227 \\ -0.239 & 1.041 & 0.261 & 0.005 \\ -0.003 & 0.261 & 1.136 & 0.283 \\ 0.227 & 0.005 & 0.283 & 1.127 \end{pmatrix} \quad \hat{\Sigma} = \begin{pmatrix} 0.912 & -0.239 & 0.295 & 0.002 \\ -0.239 & 1.041 & -0.322 & 0.261 \\ 0.295 & -0.322 & 0.928 & -0.053 \\ 0.002 & 0.261 & -0.053 & 0.873 \end{pmatrix}$$

$$\Sigma_{[14]} = \begin{pmatrix} 0.912 & 0.227 \\ 0.227 & 1.127 \end{pmatrix} \qquad \hat{\Sigma}_{[14]} = \begin{pmatrix} 0.912 & 0.002 \\ 0.002 & 0.873 \end{pmatrix}$$

$$(\Sigma_{[14]})^{-1} = \begin{pmatrix} 1.154 & -0.232 \\ -0.232 & 0.933 \end{pmatrix} \qquad (\hat{\Sigma}_{[14]})^{-1} = \begin{pmatrix} 1.096 & -0.003 \\ -0.003 & 1.145 \end{pmatrix}$$

$$J = \begin{pmatrix} 1.175 & 0.285 & \mathbf{0} & -0.020 \\ 0.285 & 1.088 & -0.250 & \mathbf{0} \\ \mathbf{0} & -0.250 & 1.000 & -0.250 \\ -0.020 & \mathbf{0} & -0.250 & 1.211 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 0.912 & -0.253 & -0.062 & 0.002 \\ -0.253 & 1.048 & 0.275 & 0.052 \\ -0.062 & 0.275 & 1.126 & 0.231 \\ 0.002 & 0.052 & 0.231 & 0.873 \end{pmatrix}$$

Gaussian IPF

Gaussian distribution:

$$\mathcal{N}(x\,;\,\mu,\Sigma) = \mathcal{N}^{-1}(x\,;\,h,J)$$

where $J = \Sigma^{-1}$

$h = \mu\,\Sigma^{-1}$

- Initialize; choose a factor $f(x_i,x_j)$
- Update to match empirical marginal

$$J = \begin{pmatrix} 1.167 & 0.268 & 0 & -0.008 \\ 0.268 & 1.137 & 0.373 & 0 \\ 0 & 0.373 & 1.210 & 0.066 \\ -0.008 & 0 & 0.066 & 1.149 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.912 & -0.239 & 0.074 & 0.002 \\ -0.239 & 1.041 & -0.322 & 0.016 \\ 0.074 & -0.322 & 0.928 & -0.053 \\ 0.002 & 0.016 & -0.053 & 0.873 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 0.912 & -0.239 & 0.295 & 0.002 \\ -0.239 & 1.041 & -0.322 & 0.261 \\ 0.295 & -0.322 & 0.928 & -0.053 \\ 0.002 & 0.261 & -0.053 & 0.873 \end{pmatrix}$$

- Iterate through all edges until convergence
- At convergence:
  - Non-edges remain zero (sparse; conditional independence)
  - On all edges, features match their empirical expectations
  - On non-edges, empirical expectations may not match (model mismatch)