- Gaussian Distribution

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-u)^T \Sigma^{-1} (x-u)\right] \qquad \Sigma = E[XX^T] - E[X]\,E[X]^T$$

$J = \Sigma^{-1}$ is the information matrix / precision matrix

$p(x) \propto \exp\left[-\frac{1}{2} x^T J x + (Ju)^T x\right]$ is the information form

$h = Ju$ is called the potential vector

Let X, Y have a joint normal distribution

$$p(X,Y) \sim \mathcal{N}\left(\begin{bmatrix} u_X \\ u_Y \end{bmatrix} ; \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$

the marginal distribution over Y is a normal distribution $\mathcal{N}(u_Y ; \Sigma_{YY})$

Let $X = X_1, \dots, X_n$ have a joint distribution $\mathcal{N}(u, \Sigma)$.

$$X_i \perp X_j \iff \Sigma_{ij} = 0$$

$$(X_i \perp X_j \mid X - \{X_i, X_j\}) \iff J_{ij} = 0 \qquad (J = \Sigma^{-1})$$

can view the information matrix as directly defining a minimal I-map Markov network for p.
whereby nonzero entries corresponds to edges in the network

- Gaussian Bayesian Network

A Gaussian Bayesian Network is a Bayesian network all of whose variables are continuous, and all CPDs are linear Gaussian

Let Y be a linear Gaussian of its parents $X_1, \dots, X_k$

$$p(Y|X) = \mathcal{N}(\beta_0 + \beta^T X ; \sigma^2)$$

Assume that $X_1 \cdots X_k$ are jointly Gaussian $\mathcal{N}(u, \Sigma)$

The distribution of Y is a normal $p(Y) = \mathcal{N}(u_Y, \sigma_Y^2)$ $\qquad u_Y = \beta_0 + \beta^T u \qquad \sigma_Y^2 = \sigma^2 + \beta^T \Sigma \beta$

The joint distribution over $\{X_i, Y\}$ is a normal where $\quad Cov[X_i ; Y] = \sum_{j=1}^{k} \beta_j \Sigma_{i,j}$

if $\mathcal{B}$ is a linear Gaussian Bayesian network, then it defines a Gaussian joint distribution let $\{X, Y\}$ have a joint normal distribution $p(X,Y) = \mathcal{N}\left(\begin{bmatrix}u_X\\ u_Y\end{bmatrix}; \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY}\\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$

then the conditional probability is a linear Gaussian

$$p(Y|X) = \mathcal{N}(\beta_0 + \beta^T X ; \sigma^2)$$

$$\beta_0 = u_Y - \Sigma_{YX}\Sigma_{XX}^{-1} u_X \qquad u_{Y|X} = u_Y - \Sigma_{YX}\Sigma_{XX}^{-1}(X - u_X)$$

$$\beta = \Sigma_{XX}^{-1}\Sigma_{YX}$$

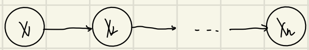
$$\sigma^2 = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$$

let $X = \{X_1, \dots, X_n\}$ and p be a joint Gaussian distribution over X.

given any ordering $X_1 \cdots X_n$ over X, can construct a Bayesian Network graph G and a Bayesian network B over G such that

- $Pa(X_i) \subseteq \{X_1, \cdots X_{i-1}\}$
- CPD of X_i in B is a linear Gaussian of its parents
- G is a minimal I-map for p

eg.

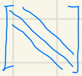


X_{i+1} is a linear Gaussian of X_i

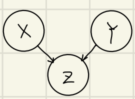
each pair of (X_i, X_j) is marginally correlated, the covariance is dense.

each pair of (X_i, X_j) is conditionally independent given other variables

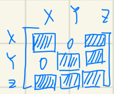
$\therefore$ the information matrix is tri-diagonal

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

eg.



Z is linear Gaussian of X and Y



X and Y are $\begin{cases} \text{marginally independent}, \therefore \text{Covariance matrix is sparse} \\ \text{conditionally dependent}, \therefore \text{information matrix is dense} \end{cases}$

· Gaussian Markov Random Fields

$$\mathcal{N}(u;\Sigma) = -\frac{1}{2}(x-u)^T\Sigma^{-1}(x-u)$$
$$= -\frac{1}{2}(x^TJx - 2x^TJu + u^TJu)$$
$$p(x) \propto \exp\left[-\frac{1}{2}x^TJx + (Ju)^Tx\right]$$

$$[x^T][J][x] + [h^T][x]$$

$$-\frac{1}{2}x^TJx + (Ju)^Tx$$
$$= \sum_i -\frac{1}{2}x_iJ_{ii}x_i + h_ix_i + \sum_{i\neq j} -\frac{1}{2}x_iJ_{ij}x_j$$

the information form immediately induces a pairwise Markov Network

the node potentials are derived from the potential vector and diagonal elements of information matrix $\left(-\frac{1}{2}x_iJ_{ii}x_i + h_ix_i\right)$

the edge potentials are derived from the off-diagonal elements of the information matrix $\left(-\frac{1}{2}x_iJ_{ij}x_j\right)$

when $J_{ij}=0$ there is no edge between X_i and X_j in the model. corresponding directly to the independence assumption of the Markov network.

Any Gaussian distribution can be represented as a pairwise Markov network with quadratic node potential and edge potential. often called a Gaussian Markov random field (GMRF)

For any pairwise Markov network with quadratic node and edge potential

$$\epsilon_i(x_i) = d_0^i + d_1^ix_i + d_2^ix_i^2$$
$$\epsilon_{ij}(x_i,x_j) = a_{00}^{ij} + a_{01}^{ij}x_i + a_{10}^{ij}x_j + a_{11}^{ij}x_ix_j + a_{02}^{ij}x_i^2 + a_{20}^{ij}x_j^2$$

can reformulate any such set of potentials in the log-quadratic form

$$p'(x) = \exp\left(-\frac{1}{2}x^TJx + h^Tx\right)$$

A Markov Network defines a valid Gaussian density $\Longleftrightarrow$ J is a pd matrix

A Quadratic Markov Random Field parametrized by J is diagonally dominant if

$$\sum_{j\neq i}|J_{ij}| < J_{ii} \quad \forall i$$

diagonally dominant $\Longrightarrow$ valid Gaussian MRF

A quadratic MRF parameterized as this is pairwise normalizable if

$$d_2^i > 0 \qquad \begin{bmatrix} a_{02}^{ij} & a_{11}^{ij}/2 \\ a_{11}^{ij}/2 & a_{20}^{ij} \end{bmatrix} \succeq 0$$

pairwise normalizable $\Rightarrow$ valid Gaussian MRF