Introduction to graphical models

Nicholas Eriksson

Department of Statistics
Stanford University

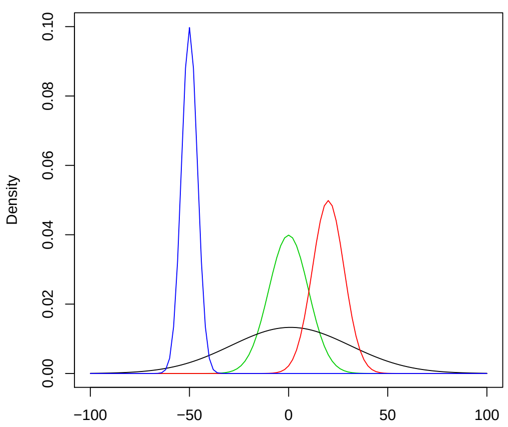
July 2, 2007

Outline

Statistical models

Definition

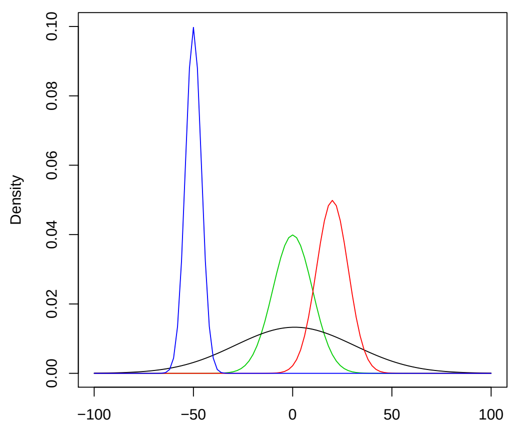
A statistical model is a “family” of probability distributions.



Example

The normal distribution: $X \sim N(\mu, \sigma)$

$$P(X = x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



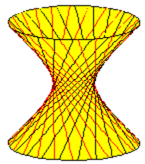
Example

Two independent binary random variables X, Y

Independence means: $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Set $\Pr(X = 0) = \alpha, \qquad \Pr(Y = 0) = \beta.$

$$\Pr(X, Y) \quad : \quad \begin{array}{c c} & \begin{array}{cc} X = 0 & X = 1 \end{array} \\ \begin{array}{c} Y = 0 \\ Y = 1 \end{array} & \begin{pmatrix} \alpha\beta & (1-\alpha)\beta \\ \alpha(1-\beta) & (1-\alpha)(1-\beta) \end{pmatrix} \end{array}$$

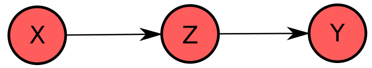


A simple graphical model

Three binary random variables:

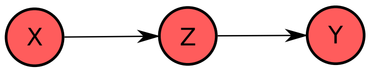
- X = does the person smoke?
- Z = do his lungs have tar?
- Y = does he have lung cancer?

How do these variables interact with each other?



Joint probability distribution

The graphical model



implies a factorization

$$\Pr(X, Z, Y) = \Pr(X)\Pr(Z|X)\Pr(Y|Z)$$

5 parameters:

$$\theta^X = \Pr(X = 0) \quad \begin{array}{l} \theta_0^Z = \Pr(Z = 0|X = 0) \\ \theta_1^Z = \Pr(Z = 0|X = 1) \end{array} \quad \begin{array}{l} \theta_0^Y = \Pr(Y = 0|Z = 0) \\ \theta_1^Y = \Pr(Y = 0|Z = 1) \end{array}$$

Relative risk

How much does the risk of lung cancer increase if a person smokes?
One answer: the (log) relative risk:

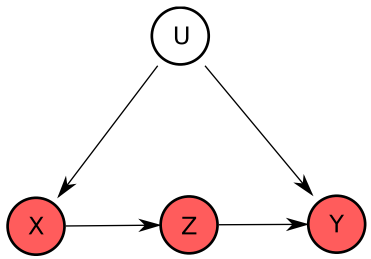
$$RR = \log\left(\frac{\Pr(Y = 1 \mid X = 1)}{\Pr(Y = 1 \mid X = 0)}\right)$$

If we believe the model, we can estimate the risk by estimating parameters:

$$\hat{RR} = \log \frac{(1-\hat\theta^X)\hat\theta_1^Z(1-\hat\theta_0^Y) + (1-\hat\theta^X)(1-\hat\theta_1^Z)(1-\hat\theta_1^Y)}{\hat\theta^X\hat\theta_0^Z(1-\hat\theta_0^Y) + \hat\theta^X(1-\hat\theta_0^Z)(1-\hat\theta_1^Y)}$$

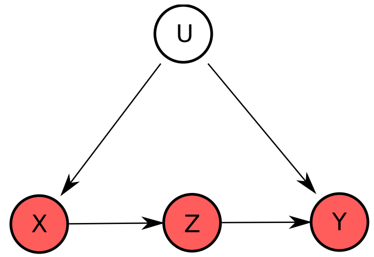
Hidden variables

Confounding factors: some (unknown) factors might make a person more inclined to smoke and also to develop cancer.



How could we compute the relative risk without being able to see U?

Joint probabilities



$$\Pr(U, X, Z, Y) = \Pr(U)\Pr(X|U)\Pr(Z|X)\Pr(Y|U, Z)$$

Relative risk

$$\Pr(U, X, Z, Y) = \Pr(U)\Pr(X|U)\Pr(Z|X)\Pr(Y|U, Z)$$

9 parameters:

θ^U	θ^X	θ^Z	θ^Y
$\Pr(U = 0)$	$\Pr(X = 0 \mid U = 0)$	$\Pr(Z = 0 \mid X = 0)$	$\Pr(Y = 0 \mid U = 0, Z = 0)$
	$\Pr(X = 0 \mid U = 1)$	$\Pr(Z = 0 \mid X = 1)$	$\Pr(Y = 0 \mid U = 0, Z = 1)$
			$\Pr(Y = 0 \mid U = 1, Z = 0)$
			$\Pr(Y = 0 \mid U = 1, Z = 1)$

But we can't observe U, so how do we estimate θ?

Graphical models

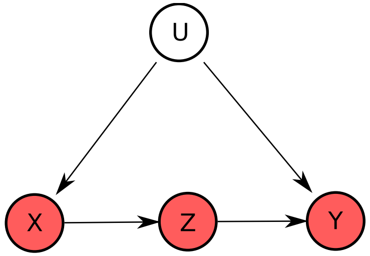
Definition of a graphical model

Let G be a directed, simple graph with vertex set $V = \{X_1, \ldots, X_n\}$, where each X_i is a random variable. The graphical model $\mathcal{M}_G$ is given by the set of all joint probability distributions on X that factor according to:

$$\Pr(X_1, \ldots, X_n) = \prod_i \Pr(X_i \mid \mathrm{pa}(X_i))$$

where $\mathrm{pa}(X_i)$ is the set of parents of X_i in G.

Example



$$\Pr(U, X, Z, Y) = \Pr(U)\Pr(X|U)\Pr(Z|X)\Pr(Y|U, Z)$$

The goal

1. Given some variables, learn how they depend on each other.
2. Using this structure, make deductions about the system.

Likelihood

- Random variables $\mathbf{X} = (X_1, \ldots, X_n)$.
- Statistical model: $\Pr_\theta(\mathbf{X} = \mathbf{x})$, parameters $\theta = (\theta_1, \ldots, \theta_k)$.
- Data: counts $u_\mathbf{x}$
- log-likelihood of the data given the model:

$$l(\theta, u) = \sum_{\mathbf{x}} u_\mathbf{x} \cdot \log \Pr_\theta(\mathbf{X} = \mathbf{x})$$

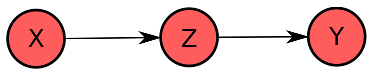
- Maximum likelihood: find θ that maximizes $l(\theta, u)$.

Maximization

Solve the system of equations

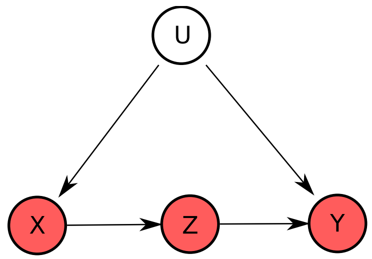
$$0 = \frac{\partial l(\theta, u)}{\partial \theta_i} = \sum_{\mathbf{x}} u_\mathbf{x} \frac{\frac{\partial}{\partial \theta_i} \Pr_\theta(\mathbf{X} = \mathbf{x})}{\Pr_\theta(\mathbf{X} = \mathbf{x})}$$

Observed likelihood



$$\Pr(X, Y, Z) = \Pr(X)\Pr(Z|X)\Pr(Y|Z)$$

Hidden likelihood



$$\Pr(X, Z, Y) = \Pr(U = 0)\Pr(X|U = 0)\Pr(Z|X)\Pr(Y|U = 0, Z) + \Pr(U = 1)\Pr(X|U = 1)\Pr(Z|X)\Pr(Y|U = 1, Z)$$

If we observed U, the likelihood equations would be easy to solve again... so we will just guess the value of U.

The EM algorithm

- Statistical model with parameters θ, on random variables $\mathbf{X}_{\text{obs}} = (X_1, \dots, X_n)$ and $\mathbf{X}_{\text{hid}} = (X_{n+1}, \dots, X_m)$
- Observed data $u_{\mathbf{x}_{\text{obs}}}$. We will guess the hidden data $u_{\mathbf{x}_{\text{hid}}, \mathbf{x}_{\text{obs}}}$.

Algorithm

Two steps:

E-step: Given a parameter estimate $\hat{\theta}$, calculate $\hat{u}_{\mathbf{x}_{\text{hid}}, \mathbf{x}_{\text{obs}}} = \mathbb{E}_{\hat{\theta}}(u_{\mathbf{x}_{\text{hid}}, \mathbf{x}_{\text{obs}}} \mid u_{\mathbf{x}_{\text{obs}}})$.

M-step: Given $\hat{u}_{\mathbf{x}_{\text{hid}}, \mathbf{x}_{\text{obs}}}$, calculate parameter estimates $\hat{\theta}$.

Begin by guessing parameters $\hat{\theta}$. Repeat the two steps until the likelihood converges.

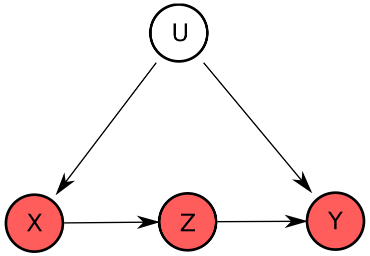
Local maximum

Theorem

The EM algorithm converges to a local maximum of the likelihood function.

To try to find a global maximum: repeat with random starting points, or use a heuristic to approximate the global maximum before starting.

Example



$$\Pr(U, X, Z, Y) = \Pr(U)\Pr(X|U)\Pr(Z|X)\Pr(Y|U, Z)$$

θ^U	θ^X	θ^Z	θ^Y
$\Pr(U = 0)$	$\Pr(X = 0\|U = 0)$	$\Pr(Z = 0\|X = 0)$	$\Pr(Y = 0\|U = 0, Z = 0)$
	$\Pr(X = 0\|U = 1)$	$\Pr(Z = 0\|X = 1)$	$\Pr(Y = 0\|U = 0, Z = 1)$
			$\Pr(Y = 0\|U = 1, Z = 0)$
			$\Pr(Y = 0\|U = 1, Z = 1)$