Quantifying Uncertainty

Sai Ravela

M. I. T

2012

Hierarchical Bayesian Models

So far: Relationships at a single level of abstraction

$$y \sim f(x)$$

Hierarchical Bayes -representation and reasoning at multiple levels of abstraction

Motivating Example: Generalized Linear Model

$$g(\mu_i) = \beta_0 + \beta_1 x_i^{(1)} + \ldots + \beta_1 x_i^{(n)}$$

$$y_i \sim P(\mu_i) \qquad i = 1, \ldots, N$$

$$P\left(\beta_0, \ldots, \beta_n | \{y_i\}, \{x_i\}\right)$$

$\Rightarrow$ Estimate the posterior distribution of model parameters

A Bayesian Perspective

Lets write:

$$
\begin{aligned}
&P(\beta_0, \ldots, \beta_n, |\{y_i\}, \{x_i\}) \\
&\quad = P\left(\underline{\beta}, |\{y_i\}, \{x_i\}\right) \\
&\quad = P(\{y_i\}|\{x_i\}, \underline{\beta}) P(\{\underline{\beta}, \{x_i\}) \\
&\quad = \prod_{i=1}^{N} \underbrace{P(y_i|x_i, \underline{\beta})}_{Likelihood} \underbrace{P(\underline{\beta}|x_i)}_{Parameters} \underbrace{P(x_i)}_{Predictors}
\end{aligned}
$$

glmfit in matlab, can find the MLE (not MAP) for parameters of Generalized linear Models.

Hyper parameters

$$\underline{\beta} \sim P(\cdot, \underline{\hat{\alpha}})$$

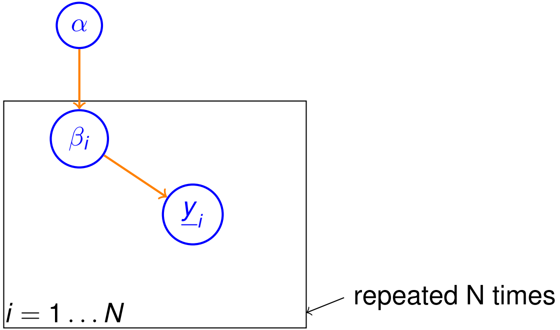
$$\prod_i P(y_i|x_i, \underline{\beta})P(\underline{\beta})$$

$$= \sum_{\underline{\hat{\alpha}}} \prod_i \underbrace{P(y_i|x_i, \underline{\beta})}_{data} \underbrace{P(\underline{\beta}|\underline{\hat{\alpha}})}_{parameter} \underbrace{P(\underline{\hat{\alpha}})}_{\substack{Hyper\\parameter}}$$

Integral may be hard to evaluate; use MCMC

Graphically

$$P(\beta_i,\alpha|\{\underline{y}_i\}) \propto \prod_{i=1}^{N} P(\underline{y}_i|\beta_i)P(\beta_i|\alpha)P(\alpha)$$



So, how to calculate posterior
-MCMC/Gibbs (very popular)

Note

This is a generic mechanism to model dependancies

Examples

Speech, Language, Climate,. . .
$\Rightarrow$ An explosion of applications

Example

(Elsner & Jagger 04)

$$y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 CT1 + \beta_2 NAOI + \beta_3 CTI \times NAOI$$

$$\underline{\beta} \sim N(\underline{\mu}, \Sigma^{-1})$$

Also read "Regression Machines" for Generalized Linear Models

Poisson Distribution

$$Pois(y;\lambda) = \frac{\lambda^y}{y!}e^{-\lambda} \quad (1)$$

Mean: λ
Variance: λ

Model Seasonal effects on hurricane
“counts”

CTI: Cold Tongue Index (Average SST anomalies over 6N-6S; 180-90W)

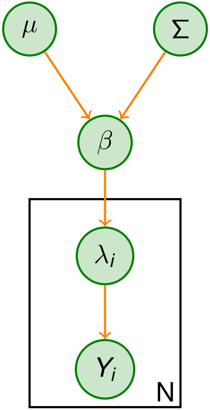
NAO Index: SLP over Gibralter

May-june: NAOI

Aug-October: CTI

Hierrachical Bayesian Model Specification

Solve using Gibbs Sampling



(Software is available, e.g. BUGS, to solve problems like this).

Constructing priors

A. Conjugate Priors: The Gamma Distribution is a conjugate prior of the Poisson Distribution; so that is one route.
B. Non-informative Prior (flat)
C. Bootstrap-Prior: Use a portion of the data to estimate parameters by MLE.

Other parameter estimates

Frequentist$\Rightarrow$ MLE
(e.g. GLM)

Contd.

Bootstrap:
Sample sets k=1...M (with replacement) from 1851-1899 Use an

MLE Estimate of $\underline{\beta}^{(k)}$

$$y_i \sim Pois(\lambda_i)$$
$$\log(\lambda_i) = \beta_0^{(k)} + \beta_1^{(k)} CTI_i + \beta_2^{(k)} NAOI_i + \beta_3^{(k)} CTI_i \times NAOI_i$$

Estimate $\underline{\beta}^{(k)}$ for each sample

$$\mu = \overline{\underline{\beta}} \qquad \Sigma^{-1} = \frac{\sum_n \underline{\beta}^{(k)} \underline{\beta}^{(k)T} - \mu\mu^T}{M-1}$$

Estimated Parameter Uncertainties

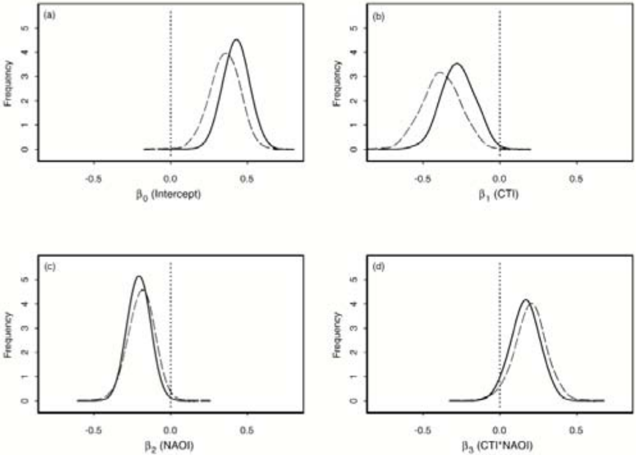


FIG. 6. Posterior distributions of the regression coefficients using the bootstrap (solid) and noninformative (dashed) priors as part of the Bayesian model of U.S. hurricane activity. (a) Intercept term, (b) ENSO term, (c) NAO term, and (d) interaction term.

example

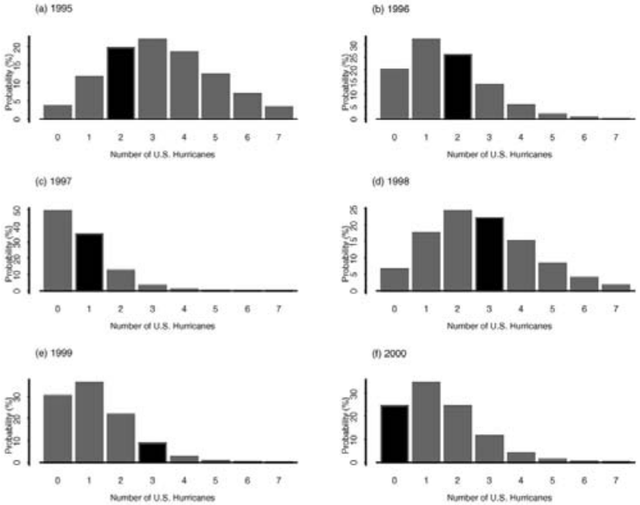
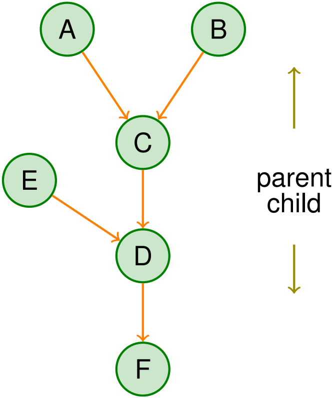


FIG. 9. (a)–(f) Predictive distributions of the number of U.S. hurricanes over the period 1995–2000. The probability of observing H number of hurricanes where $H = 0, 1, \ldots, 7$ is given on ordinate. The black bar indicates the observed number of hurricanes for that year. There were two U.S. hurricanes in 1995 (Erin and Opal).

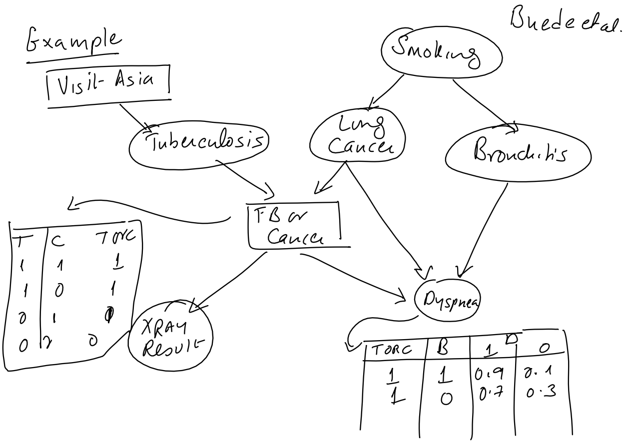
Generalizing



- a Hierarchical relationship between variables
- b All are random
- c Represented by directed acyclic graphs

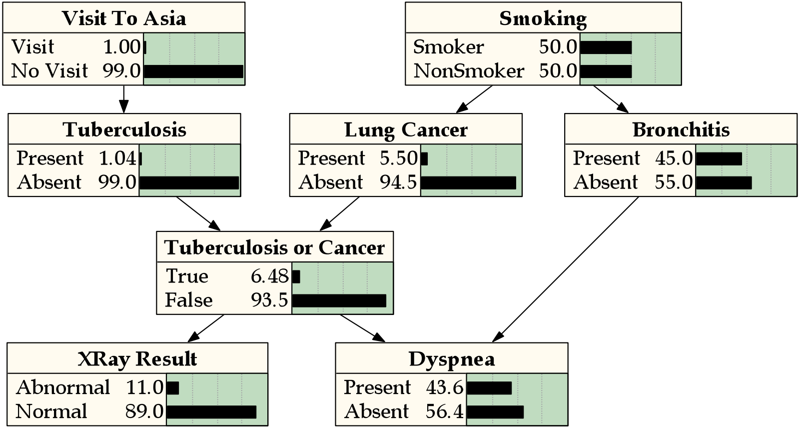
⇒ Bayesian Networks

example



Buede et al.

example


Visit To Asia
Visit 1.00
No Visit 99.0
Smoking
Smoker 50.0
NonSmoker 50.0
Tuberculosis
Present 1.04
Absent 99.0
Lung Cancer
Present 5.50
Absent 94.5
Bronchitis
Present 45.0
Absent 55.0
Tuberculosis or Cancer
True 6.48
False 93.5
XRay Result
Abnormal 11.0
Normal 89.0
Dyspnea
Present 43.6
Absent 56.4

Bayesian Networks can be used to model dependencies and assess

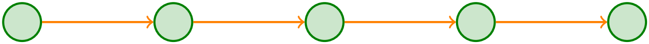
a Evidence

b Uncertainty

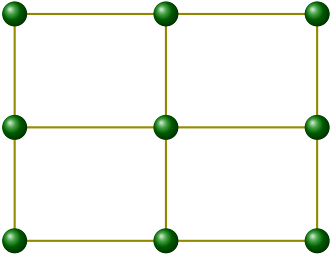
Lots of material and papers using Bayesian Networks. Ecological, Environmental and Climate applications are rapidly emerging.

Markov Networks

A markov chain is a Bayesian Network

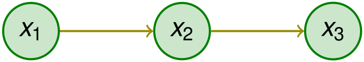


We may model “lattices” through Markov Networks



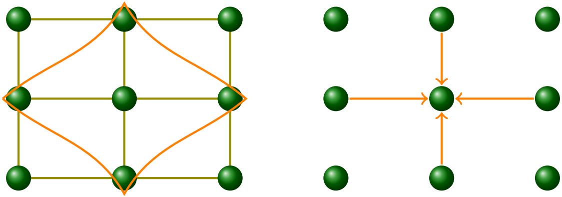
Markov random field example “ two-way interactions”

Recall again



$$
\begin{aligned}
P(x_1, x_2, x_3) &= p(x_1|x_2, x_3)p(x_2|x_3)p(x_3) \\
&= p(x_1|x_2)p(x_2|x_3)p(x_3) \\
&= p(x_3|x_2)p(x_2|x_1)p(x_1) \\
&= p(x_1|x_2)p(x_3|x_2)p(x_2)
\end{aligned}
$$

The joint and conditional views



$$P(x) = \prod_{ij} \psi_{ij}(x_{ij}, N_c(x_{ij}))$$

$$P(x_i | \underline{x} \backslash x(i,j)) = p(x_{ij} | N(x_{ij}))$$

For Bayesian Networks

$$p(x) = \prod_i p(x_i|pa(x_i))$$

pa -parent

For Markov Networks

$$p(x) = \prod_i p(x_i) \prod_{ij} p(x_i, x_j)$$

Inference on

- Markov Networks
- Bayesian(Belief) Networks

Via

- Graphical Models (see lecture notes).

12.S990 Quantifying Uncertainty
Fall 2012