Date: Thursday, March 7, 2019

Prior for Normal Distribution

$$x_j \sim N(\mu, \sigma)$$

i) Do we want a prior on σ?

(If no, give an estimate for this using standard deviation of data)

(If yes, see left

Standard choices:

a) $\mu \sim N(\mu_0, \sigma_0)$

μ_0 and σ_0 are fixed.
Values can use empirical Bayes or grid search or choose a value w/ motivation

E.B: $\mu_0 = \bar{x}$

$$\sigma_0 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$ Sample standard deviation

b) $\sigma^2 \sim \chi^2_{n-1}$

$\sigma \sim U(0, \text{max val})$

Gamma distribution Exponential

Beta-Binomial Model

Data: $X_j \sim \text{Binomial}(n, p)$ (Fixed)

↳ often used for rate problems (click thru, conversion, etc)

↓ Modeling count

$$p \sim \text{Beta}(\alpha, \beta)$$

$$\mathbb{E}[p] = \frac{\alpha}{\alpha+\beta}, \quad \text{var}(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

EB approoch: * first estimate $\frac{\bar{x}}{n}$ and $\frac{\hat{\sigma}^2}{n^2}$ from data $(x_1, \ldots, x_n)$

* solve above by plugging in $\mathbb{E}[p] = \frac{\bar{x}}{n}$, $\text{var}(p) = \frac{\hat{\sigma}^2}{n^2}$

The beta distribution is used to model data that lies between 0 and 1. Its pdf can have many shapes, depends on α, β. when $\alpha = \beta = 1$, the Beta(α, β) is a Unif$(0,1)$.

Rate Questions: Bernoulli vs Binomial

- In the context of individuals clicking on an ad
- Use Bernoulli(p) if the individual only sees this ad once (one interaction).
- Use Binomial(n,p) if there are n interactions w/ the add.

  (Binomial always works because its a generalization of Bernoulli).

Binomial vs Poisson

- Both for counts
- Bin(n,p) has n as maximum
- Po(λ) has no known maximum

Mean (count): $\mathbb{E}[Bin(n,p)] = np$

$\mathbb{E}[Po(\lambda)] = \lambda$

Major uses/applications of Bayesian Modeling Today

- Bayesian hypothesis testing
  - ↳ Results give probability of each hypothesis being true given the data.
  - ↳ More easily interpreted than p-values

- Classification:
  - ↳ Bayesian classifiers
  - ↳ Naive Bayes
  - ↳ Discriminant analysis

- Topic analysis
  - ↳ Latent Dirichlet Allocation

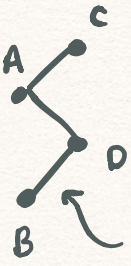
    Aim: Identify topics for a corpora of documents and calculate probability for each topic.

- Bayesian Networks (AKA Directed Acyclic Graphs DAGs)

  Idea: We have n data objects (vectors, documents, epigenes, etc) where some objects are thought to cause or heavily influence others.

  Model: probabilistic causal model where we construct a graph so that edges from node A→B implies A caused B.

Example: Data: $\{A, B, C, D\}$



edges are weighted according to degree of causality.

In DAGs, there are two main objectives

1) From multivariate data estimate the DAG/causal graph. This relies upon techniques like Gaussian graphical models, etc.
2) Once we have DAG, calculate joint probabilities using conditioning techniques based on edges in the graph.