COMPSCI 688: Probabilistic Graphical Models

Lecture 17: (Conjugate) Bayesian Inference

Dan Sheldon

Manning College of Information and Computer Sciences
University of Massachusetts Amherst

Partially based on materials by Benjamin M. Marlin (marlin@cs.umass.edu) and Justin Domke (domke@cs.umass.edu)

Conjugate Bayesian Inference

The Easy Case: Conjugacy

Some prior-likelihood pairs have a special relationship that makes computing the posterior easy

This relationship is called **conjugacy**. It means the posterior $p(\theta|x)$ will be in the same parametric family as the prior $p(\theta)$. E.g.

$$p(\theta) = \text{Beta}(\theta|a, b) \qquad \Longrightarrow \qquad p(\theta|x) = \text{Beta}(\theta|a', b')$$

We say:

- $p(\theta)$ is a conjugate prior for $p(x|\theta)$
- $p(\theta)$ and $p(x|\theta)$ are a conjugate pair

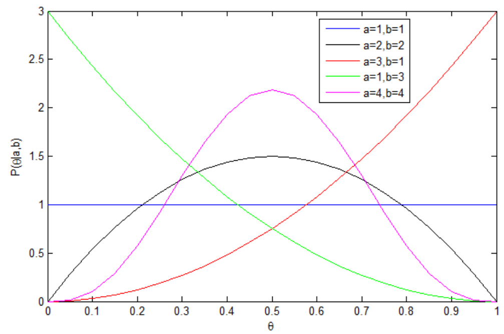
Example: Beta-Bernoulli Model

Likelihood: $p(x|\theta) = \text{Bernoulli}(x|\theta)$

Prior: $p(\theta) = \text{Beta}(\theta|a, b)$

$$\text{Beta}(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, \quad \theta \in [0, 1]$$

Beta Density



Beta Density

$$\text{Beta}(\theta|a,b) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}_{\text{normalization const.}} \underbrace{\theta^{a-1}(1-\theta)^{b-1}}_{\text{unnormalized density}}, \quad \theta \in [0,1]$$

Discuss

- Unnormalized density!
- Gamma function
  - $\Gamma(t) = \int_0^\infty z^{t-1} e^{-z} dz$
  - $\Gamma(n) = (n-1)!$ for integer n

Beta Density

$$\text{Beta}(\theta|a,b) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}_{\text{normalization const.}} \underbrace{\theta^{a-1}(1-\theta)^{b-1}}_{\text{unnormalized density}}, \quad \theta \in [0,1]$$

Question: $p(\theta) \propto \theta^2(1-\theta)^4$ on $\theta \in [0,1]$. What is normalized density?

$$\begin{aligned} p(\theta) &\propto \theta^2(1-\theta)^4 \\ &= \theta^{a-1}(1-\theta)^{b-1} \quad a=3, b=5 \\ &\propto \frac{\Gamma(8)}{\Gamma(3)\Gamma(5)}\theta^2(1-\theta)^4 \qquad \text{(normalized)} \end{aligned}$$

The point: recognize unnormalized density, get normalization constant for free

Beta-Bernoulli Posterior

Observe x. **Easy way**: drop all terms that don't involve θ

$$\begin{aligned} p(\theta|x) &= \frac{p(\theta)p(x|\theta)}{\int p(\theta')p(x|\theta')d\theta'} \\ &\propto p(\theta)p(x|\theta) \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \cdot \theta^{\mathbb{I}[x=1]}(1-\theta)^{\mathbb{I}[x=0]} \\ &\propto \theta^{a-1+\mathbb{I}[x=1]}(1-\theta)^{b-1+\mathbb{I}[x=0]} \end{aligned}$$

$$p(\theta \,|\, x) = \text{Beta}(\theta \,|\, a + \mathbb{I}[x=1], b + \mathbb{I}[x=0])$$

Result: posterior is also Beta (**conjugate**!). Add one to either a or b depending on value of x.

Beta-Bernoulli Belief Updating

Observe $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$, want to compute $p(\theta|x^{(1)}, \ldots, x^{(N)})$

By applying the simple posterior update we just saw sequentially, we get

$$p(\theta|x^{(1)}, \ldots x^{(N)}) = \text{Beta}(\theta \mid\ a + \textstyle\sum_{n=1}^{N} \mathbb{I}[x^{(n)} = 1],\ b + \sum_{n=1}^{N} \mathbb{I}[x^{(n)} = 0])$$

$$= \text{Beta}(\theta \mid\ a + \#(X = 1),\ b + \#(X = 0))$$

Simple updates based on counting

Beta-Bernoulli Belief Updating

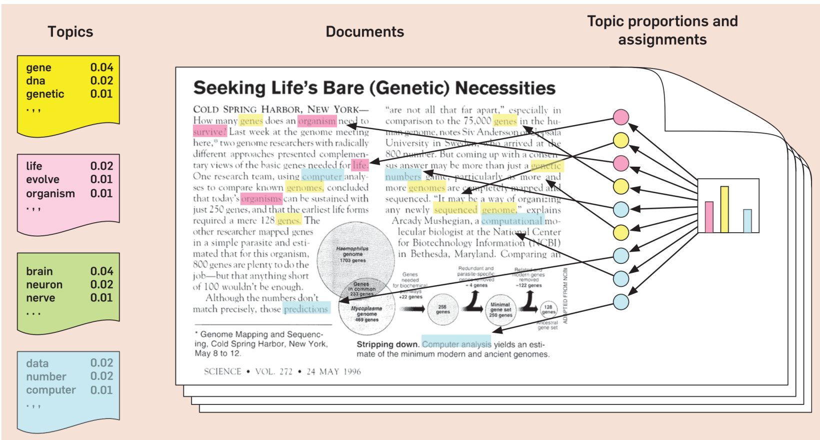
Demo

LDA

Latent Dirichlet Allocation (LDA)

Bayesian Modeling with generic inference techniques like MCMC is powerful. We can write down a generative model that we think is a good match to our data and perform inference.

LDA



Categorical and Dirichlet Distributions

- The categorical distribution is what we have been calling the multinomial distribution. For $\theta = (\theta_1, \ldots, \theta_K)$:

$$z \sim \text{Categorical}(\theta) \qquad \Longrightarrow \qquad p(z) = \prod_{k=1}^{K} \theta_k^{\mathbb{I}[z=k]}$$

- The Dirichlet is a distribution over vectors θ such that $\sum_{k=1}^{K} \theta_k = 1$ and is the conjugate prior to the multinomial:

$$\theta \sim \text{Dirichlet}(\alpha) \qquad \Longrightarrow \qquad p(\alpha) \propto \prod_{k=1}^{K} \theta_k^{\alpha}$$

LDA Model

Topic distributions:

$$\beta_k \sim \text{Dirichlet}(\eta), k = 1, \ldots, K$$

For each document d:

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

For each word position n in doc d:

$$z_{d,n} \sim \text{Cat}(\theta_d)$$
$$w_{d,n} \sim \text{Cat}(\beta_{z_n})$$

Plate Notation

We can draw the same thing compactly in plate notation to indicate repetition

Computing the Posterior

The posterior in this model looks like this:

We could sample from this unnormalized distribution using MCMC.

Example

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

Example

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.