Gibbs Sampling

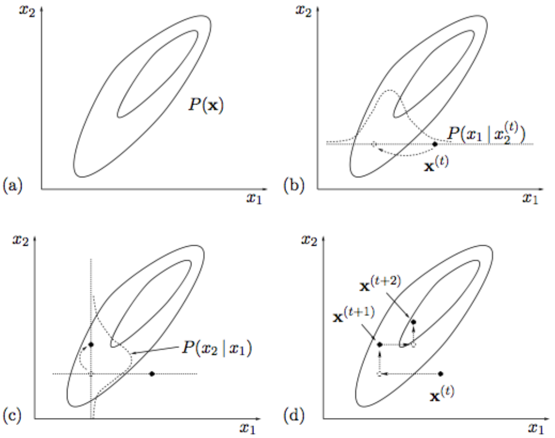
Samuel Brody

April 2009

What's it for?

- a complex problem, containing many variables
- calculating the joint probability is hard
- calculating conditional probabilities is easy

Example



from David MacKay's *Information Theory, Inference, and Learning Algorithms*

- states are points in the plane: $x = \{x_1, x_2\}$
- calculating the probability of a state is hard

$$p(x_1, x_2) =$$

$$\frac{n!}{(n-x_1)!x_1!} x_2^{x_1+\alpha-1}(1-x_2)^{n-x_1+\beta-1}$$

- calculating conditional probabilities is easy:

$$x_1 | x_2 \sim B(n, x_2)$$

$$x_2 | x_1 \sim Beta(x_1 + \alpha, n - x_1 + \beta)$$

Gibbs Sampling

- start from a random state $X_0 = \left\{X_0^{(1)}, \ldots, X_0^{(n)}\right\}$ (values of all variables)
- one by one, change each variable *i* according to the conditional distribution:

$$P\left(X^{(i)} | X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(n)}\right) = P\left(X^{(i)} | X^{(-i)}\right)$$

- repeat until convergeance

Recap - Metropolis-Hastings

- pick any transition distribution $Q_{x \to y}$
- start from a random state x_0
- in each step, sample a potential new state x^*_{i+1} from $Q_{x_i \to ?}$
- if $\frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} > 1$, move to the new (more stable) state:

  $$x_{i+1} = x^*_{i+1} \text{ w.p. } 1$$

- otherwise, move to the new state with probability $\frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}}$:

  $$\begin{aligned} x_{i+1} &= x^*_{i+1} && \text{w.p. } \frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} \\ x_{i+1} &= x_i && \text{w.p. } 1 - \frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} \end{aligned}$$

Gibbs as Metropolis-Hastings

- The Gibbs process can be viewed as a special case of the Metropolis-Hastings algorithm
- Transitions are between very similar states (only change one variable)
- Transitions are always accepted:

$$Q_{x \to y} = P(y|x) = P(y^{(i)}|x^{(-i)})$$

$$x^{(-i)} = y^{(-i)}$$

$$\frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} = \frac{P(y) \cdot Q_{y \to x}}{P(x) \cdot Q_{x \to y}} = \frac{P(y) \cdot P(x^{(i)}|y^{(-i)})}{P(x) \cdot P(y^{(i)}|x^{(-i)})} = \frac{P(y^{(i)}|y^{(-i)}) \cdot P(y^{(-i)}) \cdot P(x^{(i)}|y^{(-i)})}{P(x^{(i)}|x^{(-i)}) \cdot P(x^{(-i)}) \cdot P(y^{(i)}|x^{(-i)})} = 1$$

Gibbs as EM

Gibbs has several similarities with EM:

- iterative
- use conditional probabilities
- converge to the joint probability

The important differences:

- use randomized sampling, not most likely assignment
- one variable at a time - allows more than two sets of variables
- more flexible formulation

Local Multiple Alignment (Lawrence et al, 2005)

The Model

- a collection of protein sequences
- the sequences share a mutually similar segment

We want to find:

- the composition of the shared segment (the probability of each AA in each position)
- the location of the shared segment in each sequence a_s

A

Sigma-37	223	IIDLTYIQNK	SQKETGDILGISQMHVSR	LQRKAVKKLR	240	A25944	
SpoIIIC	94	RFGLDLKKEK	TQREIAKELGISRSYVSR	IEKRALMKMF	111	A28627	
NahR	22	VVFNQLLVDR	RVSITAENLGLTQPAVSN	ALKRLRTSLQ	39	A32837	
Antennapedia	326	FHFNRYLTRR	RRIEIAHALCLTERQIKI	WFQNRRMKWK	343	A23450	
NtrC (Brady.)	449	LTAALAATRG	NQIRAADLLGLNRNTLRK	KIRDLDIQVY	466	B26499	
DicA	22	IRYRRKNLKH	TQRSLAKALKISHVSVSQ	WERGDSEPTG	39	B24328	(BVECDA)
MerD	5	MNAY	TVSRLALDAGVSVHIVRD	YLLRGLLRPV	22	C29010	
Fis	73	LDMVMQYTRG	NQTRAALMMGINRGTLRK	KLKKYGMN	90	A32142	(DNECFS)
MAT a1	99	FRRKQSLNSK	EKEEVAKKCGITPLQVRV	WFINKRMRSK	116	A90983	(JEBY1)
Lambda cII	25	SALLNKIAML	GTEKTAEAVGVDKSQISR	WKRDWIPKFS	42	A03579	(QCBP2L)
Crp (CAP)	169	THPDGMQIKI	TRQEIGQIVGCSRETVGR	ILKMLEDQNL	186	A03553	(QRECC)
Lambda Cro	15	ITLKDYAMRF	GQTKTAKDLGVYQSAINK	AIHAGRKIFL	32	A03577	(RCBPL)
P22 Cro	12	YKKDVIDHFG	TQRAVAKALGISDAAVSQ	WKEVIPEKDA	29	A25867	(RGBP22)
AraC	196	ISDHLADSNF	DIASVAQHVCLSPSRLSH	LFRQQLGISV	213	A03554	(RGECA)
Fnr	196	FSPREFRLTM	TRGDIGNYLGLTVETISR	LLGRFQKSGM	213	A03552	(RGECF)
HtpR	252	ARWLDEDNKS	TLQELADRYGVSAERVRQ	LEKNAMKKLR	269	A00700	(RGECH)
NtrC (K.a.)	444	LTTALRHTQG	HKQEAARLLGWGRNTLTR	KLKELGME	461	A03564	(RGKBCP)
CytR	11	MKAKKQETAA	TMKDVALKAKVSTATVSR	ALMNPDKVSQ	28	A24963	(RPECCT)
DeoR	23	LQELKRSDKL	HLKDAAALLGVSEMTIRR	DLNNHSAPVV	40	A24076	(RPECDO)
GalR	3	MA	TIKDVARLAGVSVATVSR	VINNSPKASE	20	A03559	(RPECG)
LacI	5	MKPV	TLYDVAEYAGVSYQTVSR	VVNQASHVSA	22	A03558	(RPECL)
TetR	26	LLNEVGIEGL	TTRKLAQKLGVEQPTLYW	HVKNKRALLD	43	A03576	(RPECTN)
TrpR	67	IVEELLRGEM	SQRELKNELGAGIATITR	GSNSLKAAPV	84	A03568	(RPECW)
NifA	495	LIAALEKAGW	VQAKAARLLGMTPRQVAY	RIQIMDITMP	512	S02513	
SpoIIG	205	RFGLVGEEEK	TQKDVADMMGISQSYISR	LEKRIIKRLR	222	S07337	
Pin	160	QAGRLIAAGT	PRQKVAIIYDVGVSTLYK	TFPAGDK	177	S07958	
PurR	3	MA	TIKDVAKRANVSTTTVSH	VINKTRFVAE	20	S08477	
EbgR	3	MA	TLKDIAIEAGVSLATVSR	VLNDDPTLNV	20	S09205	
LexA	27	DHISQTGMPP	TRAEIAQRLGFRSPNAAE	EHLKALARKG	44	S11945	
P22 cI	25	SSILNRIAIR	GQRKVADALGINESQISR	WKGDFIPKMG	42	B25867	(Z1BPC2)
			*****************	***			

Conditional Formulation

- if we know the shared segment,
  $\Rightarrow$ we know the location probability for each sequence.
- if we know location of the segment in each sequence
  $\Rightarrow$ we know the composition of the segment.

Gibbs Inference

- randomly initialize the locations in each sequence
- randomly choose a single sequence z
- calculate the probability for the segment in each location $p(a_z = i)$
- sample a new location a'_z from this distribution

B

	Position in site																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Arg	94	222	265	137	9	9	137	137	9	9	9	52	222	94	94	9	265	606
Lys	9	133	442	380	9	71	380	194	9	133	9	9	71	9	9	9	71	256
Glu	53	9	96	401	9	9	140	140	9	9	9	53	140	140	9	9	9	53
Asp	67	9	9	473	9	9	299	125	9	67	9	67	67	9	9	9	9	67
Gln	9	600	224	9	9	9	224	9	9	9	9	9	278	63	278	9	9	170
His	240	9	9	9	9	9	125	125	9	9	9	9	125	125	125	9	9	240
Asn	168	9	9	9	9	9	168	89	9	89	9	248	9	168	89	9	89	89
Ser	117	9	117	117	9	9	9	9	9	9	9	819	63	387	63	9	819	9
Gly	151	9	56	9	9	151	9	9	9	1141	9	151	9	56	9	9	56	9
Ala	9	9	112	43	181	901	43	181	215	9	43	9	43	181	112	43	78	9
Thr	915	130	130	9	251	9	9	9	9	9	9	311	130	70	855	9	130	9
Pro	76	9	9	9	9	9	9	9	9	9	9	9	210	210	9	9	9	9
Cys	9	9	9	9	9	9	9	9	295	581	295	9	9	9	9	9	9	9
Val	58	107	9	9	500	9	9	9	156	9	598	9	205	58	9	746	9	58
Leu	9	121	9	9	149	9	93	149	458	9	149	9	37	37	9	177	9	9
Ile	9	166	114	61	323	9	114	166	9	9	427	9	61	9	61	427	9	61
Met	9	104	9	9	9	9	9	198	198	9	104	9	9	198	9	9	9	9
Tyr	9	9	136	9	9	9	9	262	262	9	9	136	136	9	262	9	262	136
Phe	9	9	9	9	9	9	9	9	9	9	108	9	9	9	9	9	9	9
Trp	9	9	9	9	9	9	9	9	9	9	366	9	9	9	9	9	9	366

Fig. 1. Alignment and probability ratio model for the helix-turn-helix pattern common to 30 proteins (*45*). **(A)** The alignment. Columns from left to right are: sequence name; locations a_k of the left end of the common pattern in each sequence; aligned sequences, including residues flanking the 18-residue common pattern; right-end positions (a_k + 17) of the common pattern; NBRF/PIR accession number; and NBRF/PIR code name, if available. Asterisks (***) below the alignment indicate the 20-residue segment previously described on the basis of structural superpositions (*26, 27*). Almost equal values of information per parameter were given by pattern widths of 18 to 21 residues (Fig. 2): the longer widths extended to the right the 18-residue pattern shown. **(B)** Probability ratios (100 × $q_{i,j}/p_j$) for each amino acid at each position in the pattern model.

Additional Issues

Still need to deal with:

- segment length
- multiple patterns
- phase shifts (local maxima)

Topic Models - LDA (Blei et al. 2003)

The Model:

- a collection of documents
- each document is a mixture of several (latent) topics
- each word in the document comes from a single topic
- topics are shared between documents

We want to find:

- the topics t_i^d
- the word-topic distribution - $\phi(w|t)$
- the topic distribution in each document - θ_d

Example - Document

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.

Example - Topics

"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

Conditional Formulation

- if we know the topic assignments t_i^d of the words,
  $\Rightarrow$ we know the prob. topic distrib. θ & word-topic distrib. $\phi(w|t)$

- if we know θ_d & $\phi(w|t)$
  $\Rightarrow$ we know the probable assignments t_i^d

Gibbs Inference

- start with a random assignment of words to topics
- calculate θ_d and $\phi(w|t)$
- for each word, sample a new assignment t_i^d, based on $\theta_d(t) \cdot \phi(w_i|t)$

Summary

What's it for?

- estimating a complex joint probability

How?

- small changes based on conditional probabilities

Keep in mind:

- burn-in period
- local maxima for typical states