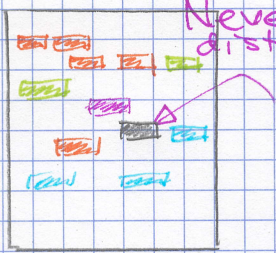
What is collapsed Gibbs sampling?

"Collapsed" Gibbs sampling for LDA

Based on special structure of LDA model, can sample *just* indicator variables z_{iw}

- No need to sample other parameters
  - corpus-wide topic vocab distributions
  - per-doc topic proportions

Often leads to much better performance because examining uncertainty in smaller space



Never draw topic vocab distributions or doc topic proportions

Randomly reassign z_{iw} based on current assignments z_{jv} of all other words in document and corpus

A worked example for LDA: Initial setup

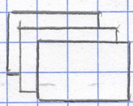
Select a document

Randomly assign topics

(one possible approach)

3	2	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

5 word document



Repeat for each doc in the corpus

Maintain local statistics

	Topic 1	Topic 2	Topic 3
Doc i	2	1	2

	Topic1	Topic2	Topic3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
~~model~~ dynamic	10	8	1
...			

Total counts from all docs

Randomly reassign topics

3	~~2~~	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

	Topic 1	Topic 2	Topic 3
Doc i	2	0 ~~1~~	2

	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	10	7 ~~8~~	1
...			

decrementing counts after removing current assignment $z_{iw} = 2$

Probability of new assignment

?
dynamic

reassign with probability

$p(z_{iw} \mid$ every other z_{jv} in corpus, words in corpus$)$

A worked example for LDA: Deriving the resampling distribution

Probability of new assignment

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Topic 1 Topic 2 Topic 3

How much doc "likes" each topic based on other assignments in doc

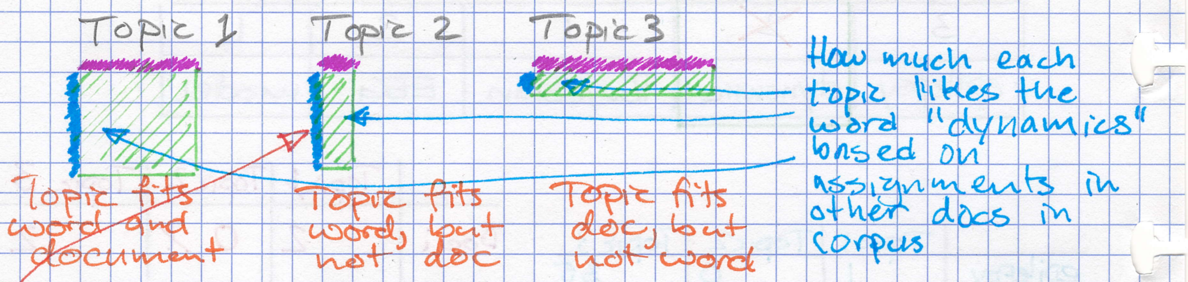
	Topic 1	Topic 2	Topic 3
Doc i	2	0	2

current assignments to topic k in doc i → $n_{ik} + \alpha$ ← smoothing param from Bayes prior

words in doc i → $N_i - 1 + K\alpha$

ignore current word

Probability of new assignment cont'd



#assignments corpus-wide of word "dynamic" to topic k

$$\frac{m_{\text{dynamic},k} + \gamma}{\sum_{w \in V} m_{w,k} + V\gamma}$$

smoothing param from Bayes prior

size of vocab

	Topic 1	Topic 2	Topic 3
dynamic	10	7	1

Topic 2 also really likes "dynamic", but in a different context... e.g. a topic on fluid dynamics

How much doc likes topic

$$\boxed{\frac{n_{ik} + \alpha}{N_i - 1 + K\alpha} \cdot \frac{m_{\text{dynamic},k} + \gamma}{\sum_{w \in V} m_{w,k} + V\gamma}}$$

How much topic likes word

Normalize this product of terms over K possible topics!

Randomly draw a new topic indicator

To draw new topic assignment (equivalently)
- roll K-sided die with these probabilities
- throw dart at these regions

Update counts

3	1	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Increase popularity of topic 1 in doc i

	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	11 ~~10~~	7	1
...			

	Topic 1	Topic 2	Topic 3
doc i	3 ~~2~~	0	2

increment counts based on new assignment of $z_{iw} = 1$

Increase popularity of "dynamic" in topic 1 (corpus-wide)

Iterate through all words/docs

3	1	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Using the output of collapsed Gibbs sampling

What to do with the collapsed samples?

From "best" sample of $\{z_{iw}\}$, can infer

1. Topics from conditional distribution...
   need corpus-wide info
2. Document "embedding"...
   need doc info only

Embedding new documents

Simple approach:

1. Fix topics based on training set collapsed sampling
2. Run uncollapsed sampler on new doc(s) only