Graphical Model : $P(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z}, \underline{w})$

(A)

Goal : find (posterior distribution over) latent variables given the observed variables

$$\Rightarrow \underbrace{P(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z} \mid \underline{w})}_{\text{main vars of interest}} = \frac{P(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z}, \underline{w})}{P(\underline{w})}$$

Graphical model gives :

$$P(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z}, \underline{w}) =$$

$$\underbrace{P(\underline{\underline{\theta}})}_{\prod_d \mathrm{Dir}(\underline{\theta}_d ; \alpha)} \underbrace{P(\underline{\underline{\phi}})}_{\prod_t \mathrm{Dir}(\underline{\phi}_t ; \beta)} \underbrace{P(\underline{z} \mid \underline{\underline{\theta}})}_{\prod_n \theta_{z_n \mid d_n}} \underbrace{P(\underline{w} \mid \underline{z}, \underline{\underline{\phi}})}_{\prod_n \phi_{w_n \mid z_n}}$$

Option 1:

- Use EM or Gibbs sampling to find $\underline{\underline{\theta}}, \underline{\underline{\phi}}$ and $\underline{z}$

$$p(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z} \mid \underline{w}) \propto p(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z}, \underline{w})$$

EM: algorithm for computing the ML (or MAP) estimate of parameters in the presence of latent variables

- $\underline{\underline{\phi}}, \underline{\underline{\theta}}$ = parameters.
- $\underline{z}$ = latent variables.

Gibbs sampling: sample $\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z}$ from $p(\underline{\underline{\theta}}, \underline{\underline{\phi}}, \underline{z} \mid \underline{w})$

Expensive: sampling a huge # parameters $DT + TW + N$

tempting to sample $\underline{\underline{\theta}}$ and $\underline{\underline{\phi}}$ less often

$\Rightarrow$ complex space, slow to converge

Option 2:

- Collapsed Gibbs sampling = can integrate out $\underline{\underline{\theta}}$ and $\underline{\underline{\phi}}$ directly and just sample $\underline{z}$

$\Rightarrow$ sample $\underline{z}$ from $p(\underline{z} \mid \underline{w})$

Goal: form $p(\underline{z} \mid \underline{w})$ and draw samples from it.

$$\Rightarrow p(\underline{z} \mid \underline{w}) = \underbrace{\frac{p(\underline{z}, \underline{w})}{p(\underline{w})}}$$

can we compute this?

<u>Numerator</u>:

$$p(\underline{z}, \underline{w}) = \int d\underline{\theta} \int d\underline{\phi} \; p(\underline{\theta}, \underline{\phi}, \underline{z}, \underline{w})$$

$$= \int d\underline{\theta} \int d\underline{\phi} \prod_d \mathrm{Dir}(\underline{\theta}_d ; \alpha) \prod_t \mathrm{Dir}(\underline{\phi}_t ; \beta) \times$$

$$\underbrace{\prod_n \theta_{z_n | d_n}}_{\prod_d \prod_t \theta_{t|d}^{N_{t|d}}} \; \underbrace{\prod_n \phi_{w_n | z_n}}_{\prod_t \prod_w \phi_{w|t}^{N_{w|t}}}$$

$$= \int d\underline{\underline{\theta}} \prod_d \left[\mathrm{Dir}(\underline{\theta}_d ; \alpha) \prod_t \theta_{t|d}^{N_{t|d}} \right] \times$$

$$\int d\underline{\underline{\phi}} \prod_t \left[\mathrm{Dir}(\underline{\phi}_t ; \beta) \prod_w \phi_{w|t}^{N_{w|t}} \right]$$

$$= \underbrace{1 \times 2}$$

$$\underbrace{\mathrm{Dir}(\underline{\theta}_d ; \alpha)}_{\text{Dirichlet prior}} \underbrace{\prod_t \theta_{t|d}^{N_{t|d}}}_{\text{Mult. likelihood.}}$$

$$1 = \prod_d \int_0^1 d\underline{\theta}_d \, \underbrace{\mathrm{Dir}(\underline{\theta}_d ; \alpha)} \prod_t \theta_{t|d}^{N_{t|d}}$$

$$= \frac{\Gamma(\alpha)}{\prod_t \Gamma(\alpha \frac{1}{T})} \prod_t \theta_{t|d}^{\alpha \frac{1}{T} - 1} \quad \delta\left(\sum_t \theta_{t|d} - 1 \right)$$

(4)

$$\therefore \prod_d \int_0^1 d\theta_{1|d} \int_0^1 d\theta_{2|d} \cdots \int_0^1 d\theta_{t|d} \frac{\Gamma(\alpha)}{\prod_t \Gamma(\alpha \frac{1}{T})} \times$$

$$\prod_t \theta_{t|d}^{\alpha\frac{1}{T}-1} \prod_t \theta_{t|d}^{N_{t|d}}$$

$$= \prod_d \frac{\Gamma(\alpha)}{\prod_t \Gamma(\alpha \frac{1}{T})} \prod_t \int_0^1 d\theta_{t|d} \underbrace{\theta_{t|d}^{\alpha\frac{1}{T}-1} \theta_{t|d}^{N_{t|d}}}_{= \theta_{t|d}^{N_{t|d}+\alpha\frac{1}{T}-1}}$$

But! $\mathrm{Dir}(\underline{\theta}_d ; \{N_{t|d} + \alpha \frac{1}{T}\})$

$$= \frac{\Gamma(N_d + \alpha)}{\prod_t \Gamma(N_{t|d} + \alpha\frac{1}{T})} \prod_t \theta_{t|d}^{N_{t|d}+\alpha\frac{1}{T}-1}.$$

and

$$\int d\underline{\theta}_d \, \mathrm{Dir}(\underline{\theta}_d ; \{N_{t|d} + \alpha\frac{1}{T}\}) = 1$$

$$\Rightarrow \frac{\Gamma(N_d+\alpha)}{\prod_t \Gamma\left(N_{t|d}+\alpha\frac{1}{T}\right)} \prod_t \int d\theta_{t|d}\, \theta_{t|d}^{N_{t|d}+\alpha\frac{1}{T}-1} = 1$$

$$\Rightarrow \prod_t \int d\theta_{t|d}\, \theta_{t|d}^{N_{t|d}+\alpha\frac{1}{T}-1}$$

$$= \frac{\prod_t \Gamma\left(N_{t|d}+\alpha\frac{1}{T}\right)}{\Gamma(N_d+\alpha)}.$$

$$\Rightarrow (1) = \prod_d \frac{\Gamma(\alpha)}{\prod_t \Gamma\left(\alpha\frac{1}{T}\right)} \frac{\prod_t \Gamma\left(N_{t|d}+\alpha\frac{1}{T}\right)}{\Gamma(N_d+\alpha)}.$$

Similarly for (2).

$\Rightarrow$ can compute numerator.

Denominator:

$$p(\underline{w}) = \sum_{\underline{z}} p(\underline{z}, \underline{w})$$

doesn't factorize nicely once $\underline{\phi}$ and $\underline{\theta}$ have been integrated out, involves T^N terms!

$$\sum_{\underline{z}} p(\underline{\theta}, \underline{\phi}, \underline{z}, \underline{w}) = \sum_{\underline{z}} \prod_d \mathrm{Dir}(\underline{\theta}_d; \alpha) \times \prod_t \mathrm{Dir}(\underline{\phi}_t; \beta) \prod_n \left[\theta_{z_n|d_n} \phi_{w_n|z_n} \right]$$

$$= \prod_d \mathrm{Dir}(\underline{\theta}_d; \alpha) \prod_t \mathrm{Dir}(\underline{\phi}_t; \beta) \times \prod_n \left[\sum_{z_n} \theta_{z_n|d_n} \phi_{w_n|z_n} \right]$$

Gibbs Sampling:

- method for sampling from distributions over 2 or more variables $\underline{x}$
- too complicated to sample from $p(\underline{x})$ directly

... but can compute / sample from $p(x_n \mid \underline{x}_{\setminus n})$

Algorithm in 2D.

1. Initialize $\underline{x}^{(0)}$ randomly and $i = 0$
2. $x_1^{(i+1)} \sim p(x_1 \mid x_2^{(i)})$
3. $x_2^{(i+1)} \sim p(x_2 \mid x_1^{(i+1)})$
4. $i = i + 1$

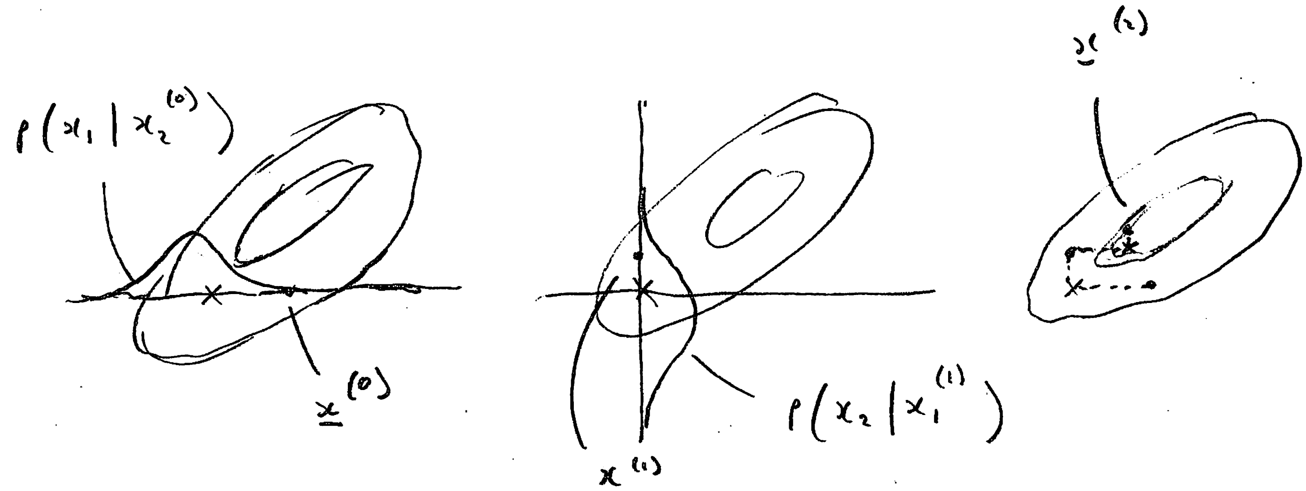
new value.

and repeat from 2

unless "converged"

Algorithm in general.

$$x_n^{(i+1)} \sim p(x_n \mid x_1^{(i+1)} \dots x_{n-1}^{(i+1)}, x_{n+1}^{(i)} \dots x_N^{(i)})$$



∴ Can we use Gibbs to sample $\underline{z}$?

Does this apply here ?

- $p(\underline{z} \mid \underline{w})$ is too complicated to compute / sample from directly ∵ of $p(\underline{w})$ ✓

What is the form of

$$p(z_n = t \mid \underline{w}, \underline{z}_{\setminus n})$$

$$= \frac{p(z_n = t, w_n \mid \underline{w}_{\setminus n}, \underline{z}_{\setminus n})}{\sum_{z_n} p(z_n, w_n \mid \underline{w}_{\setminus n}, \underline{z}_{\setminus n})}$$

sum over T terms, so if we can compute $p(z_n = t \mid w_n, \underline{w}_{\setminus n}, \underline{z}_{\setminus n})$ then we're all set !

Can we compute $p(z_n = t, w_n \mid \underline{z}_{\setminus n}, \underline{w}_{\setminus n})$?

$$= \frac{p(z_n = t, w_n, \underline{z}_{\setminus n}, \underline{w}_{\setminus n})}{p(\underline{z}_{\setminus n}, \underline{w}_{\setminus n})}$$

$p(\underline{z}, \underline{w})$

$$p(\underline{z}, \underline{w}) = \prod_d \frac{\Gamma(\alpha)}{\prod_t \Gamma(\alpha \frac{1}{T})} \frac{\prod_t \Gamma(N_{t|d} + \alpha \frac{1}{T})}{\Gamma(N_d + \alpha)} \times$$

$$\prod_t \frac{\Gamma(\beta)}{\prod_w \Gamma(\beta \frac{1}{W})} \frac{\prod_w \Gamma(N_{w|t} + \beta \frac{1}{W})}{\Gamma(N_t + \beta)}$$

$p(\underline{z}_{\setminus n}, \underline{w}_{\setminus n})$ is the same except with

$N_{t|d_n} - 1$, $N_t - 1$, $N_{d_n} - 1$, $N_{w_n|t} - 1$

$\Rightarrow$ everything cancels except ...

$$\frac{\dfrac{\Gamma\left(N_{t|d_n} + \alpha \frac{1}{T}\right)}{\Gamma\left(N_{d_n} + \alpha\right)} \; \dfrac{\Gamma\left(N_{w_n|t} + \beta \frac{1}{W}\right)}{\Gamma\left(N_t + \beta\right)}}{\dfrac{\Gamma\left(N_{t|d_n} - 1 + \alpha \frac{1}{T}\right)}{\Gamma\left(N_{d_n} - 1 + \alpha\right)} \; \dfrac{\Gamma\left(N_{w_n|t} - 1 + \beta \frac{1}{W}\right)}{\Gamma\left(N_t - 1 + \beta\right)}}$$

But! $x\,\Gamma(x) = \Gamma(x+1) \Rightarrow$

$$\frac{\Gamma(x+1)}{\Gamma(x)} = x$$

$$\Rightarrow \underbrace{\frac{N_{t|d_n} + \alpha \frac{1}{T}}{N_{d_n} + \alpha}}_{p(z_n = t \mid \underline{z}_{\setminus n})} \quad \underbrace{\frac{N_{w_n|t} + \beta \frac{1}{W}}{N_t + \beta}}_{p(w_n \mid z_n = t, \underline{z}_{\setminus n}, \underline{w}_{\setminus n})}$$

Great! So problem fits 2 criteria for Gibbs

Collapsed Gibbs for LDA:

- Repeat until converged:
  - for $n = 1 \dots N$
    - $z_n \sim p(z_n \mid \underline{w}, \underline{z}_{\setminus n})$
    - Update z_n.

<u>Convergence</u>?

- plot $p(\underline{z}, \underline{w})$ vs. iteration #