Gibbs sampling

initialize $z_1^0 \ldots z_n^0, \mu_{1:k}^0, \theta^0$

for iteration $m = 1, 2, \ldots, M$:

$$z_{1:n}^m \sim p(z_{1:n} \mid \mu_{1:k}^{m-1}, \theta^{m-1}, x_{1:n})$$

$$\mu_{1:k}^m \sim p(\mu_{1:k} \mid z_{1:n}^m, x_{1:n})$$

$$\theta^m \sim p(\pi \mid z_{1:n}^m, x_{1:n})$$

(This should feel like EM.)

This returns a set of samples

$$\{ z_{1:n}^m \; \mu_{1:k}^m \; \theta^m \}_{m=1}^M$$

Claim:

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^M \mathbb{1}(\mu_k^m \in A) = p(\mu_k \in A \mid x_{1:n})$$

for any subset A

Another way of saying this is that:

$$\lim_{M \to \infty} P_r(\mu_k^M) = p(\mu_k \mid x_{1:n})$$

More generally:

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^M f(z_{1:n}^m, \mu_{1:k}^m, \theta^m)$$

$$= \mathbb{E}[f(z_{1:n}, \mu_{1:k}, \theta) \mid x_{1:n}]$$

Posterior

Conceptually our algorithm is a <u>Markov chain</u>.

State : $S^m \triangleq (z^m_{1:n}, \mu^m_{1:k}, \theta^m)$

A Markov chain is a joint distribution of sequential random variables $S_1 \dots S_M$ which factorizes:

$$\pi(s^0, s^1, \dots s^M) = \underset{\text{"initial dist."}}{\pi(s^0)} \prod_{m=1}^{M} \underset{\text{"transition operator"}}{\pi(s^m \mid s^{m-1})}$$

If $\pi(S^m = s \mid S^{m-1} = s')$ is the same for all m then the Markov chain is <u>homogenous</u>. In this case:

$$\pi(S^m = s \mid S^{m-1} = s') \equiv \pi(s \mid s')$$

Each state $S^m \equiv (S^m_1 \dots S^m_D)$ has D components.

e.g. $S^M \equiv (z^m_1 \dots z^m_n, \mu^m_1 \dots \mu^m_k, \theta^m)$, so $D = n+k+1$

The m^{th} marginal of the chain is:

$$\pi(S^m = s) = \int \pi(S^{m-1} = s')\, \pi(S^m = s \mid S^{m-1} = s')\, ds'$$
$$\equiv \int \pi(S^{m-1} = s')\, \pi(s \mid s')\, ds'$$

A distribution $\pi^*(s)$ is a <u>stationary</u> distribution of π if:

$$\pi^*(s) = \int \pi^*(s')\, \pi(s \mid s')\, ds' \equiv \mathbb{E}_{s' \sim \pi^*}[\pi(s \mid s')]$$

We also say that π^* is invariant to π.

Detailed balance

How can we know if π^* is a stationary dist.?

A sufficient condition is if it satisfies detailed balance:

$$\pi^*(s')\,\pi(s \mid s') = \pi^*(s)\,\pi(s' \mid s)$$

To see this condition implies that π^* is stationary:

$$\int \pi^*(s')\,\pi(s \mid s')\,ds' = \int \pi^*(s)\,\pi(s' \mid s)\,ds'$$

$$= \pi^*(s)$$

Ergodicity

DB implies that π^* is a stationary dist. but not necessarily a unique one.

A Markov chain that is ergodic has a unique π^*.

$$\lim_{m \to \infty} \pi(s^m = s) = \pi^*(s), \quad \text{for any } s^0$$

A sufficient condition for ergodicity is that:

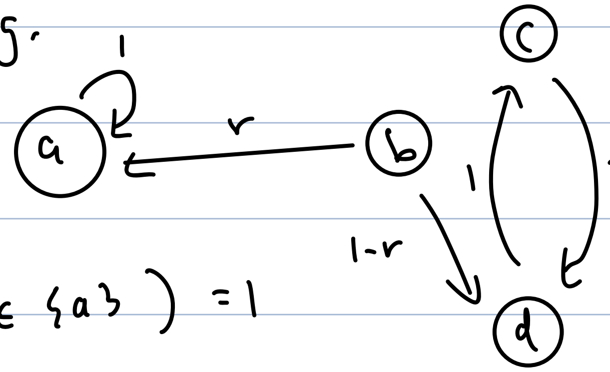
$$\pi(s \mid s') > 0 \quad \forall\, s, s'$$

More generally, a chain is ergodic if it is

1. Irreducible
2. Aperiodic

Irreducibility

e.g.

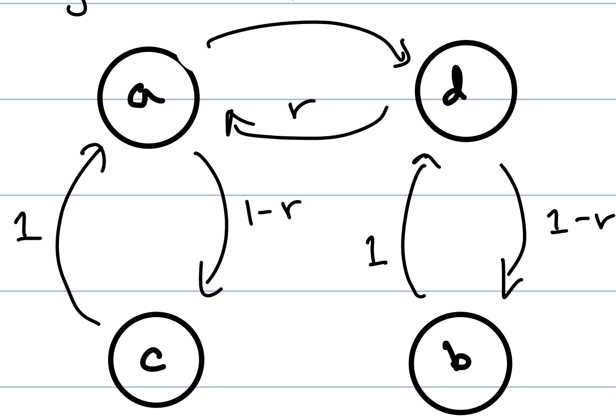


- $P(s^m \in \{a\} \mid s^{m-1} \in \{a\}) = 1$
- $P(s^m \in \{c,d\} \mid s^{m-1} \in \{c,d\}) = 1$
- $\{a\}$ and $\{c,d\}$ are recurrent classes because all nodes are reachable from all other nodes within the class
- $\{a,b,c,d\}$ is not a recurrent class, since e.g.

$$P(s^m = b \mid s^0 = c) = 0 \quad \forall\, m \geq 1$$

Periodicity

e.g.



- $P(s^m \in \{c,d\} \mid s^{m-1} \in \{a,b\}) = 1$
- $P(s^{m+1} \in \{a,b\} \mid s^m \in \{c,d\}) = 1$
- So we can partition the chain into two periodic classes $\{a,b\}$, $\{c,d\}$.

The Gibbs sampler is a Markov chain with transitions:

$$\pi_{\text{Gibbs}}(s^m \mid s^{m-1}) = p(s_1^m \mid s_2^{m-1}, \ldots s_D^{m-1}, x)$$
$$p(s_2^m \mid s_1^m, s_3^{m-1}, \ldots s_D^{m-1}, x)$$
$$\vdots$$
$$p(s_D^m \mid s_1^m \ldots s_{D-1}^m, x)$$

where $p(s_d \mid s_1 \ldots s_{d-1}, s_{d+1} \ldots s_D, x)$ is the <u>complete conditional</u> of latent variable s_d.

(These conditionals also condition on the data x)

For any coordinate d, WLOG, say that it is sampled first:

$$\pi_{\text{Gibbs}}(s_d^m = s_d) = \int p(s_d \mid s'_{\neg d}, \bar{x}) \, \pi_{\text{Gibbs}}(s_{\neg d}^{m-1} = s'_{\neg d}) \, ds'_{\neg d}$$

say that $\pi_{\text{Gibbs}}(s_{\neg d}^{m-1} = s'_{\neg d}) = p(s'_{\neg d} \mid x)$

$$= p(s_d \mid \bar{x})$$

So <u>if</u> the marginal for $s_{\neg d}^{m-1}$ is the posterior then the marginal for s_d^m is too.

It's easier to see the implication for $D = 2$:

$$\pi_{\text{Gibbs}}(s^m = (s_1, s_2))$$

$$= \int p(s_1 \mid s_2, x)\, p(s_2 \mid s_1', x)\, \pi_{\text{Gibbs}}(s^{m-1} = (s_1', s_2'))\, ds_1'\, ds_2'$$

$$= p(s_1 \mid s_2, x) \int p(s_2 \mid s_1', x)\, \pi_{\text{Gibbs}}(s_1^{m-1} = s_1')\, ds_1$$

If $\pi_{\text{Gibbs}}(s^{m-1} = (s_1', s_2')) = p(s_1', s_2' \mid x)$

$$= p(s_1, s_2 \mid x)$$

If the Gibbs chain is ergodic, then:

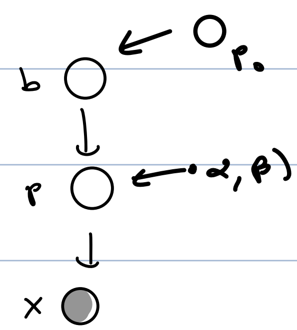
$$\lim_{m \to \infty} \pi_{\text{Gibbs}}(s^m = s) = p(s \mid \bar{x})$$

An example of a non-ergodic Gibbs chain:

$b \sim \text{Bern}(p_0)$

$$p \sim \begin{cases} \delta_0 & \text{if } b = 0 \\ \text{Beta}(\alpha, \beta) & \text{if } b = 1 \end{cases}$$

$$x \sim \begin{cases} \delta_0 & \text{if } p = 0 \\ \text{Bern}(p) & \text{if } p > 0 \end{cases}$$



$p(p = 0 \mid b = 0, \bar{x} = 0) = 1$ $\qquad$ $p(b = 0 \mid p = 0, \bar{x} = 0) = 1$

So this Gibbs sampler will never leave the state $(p = 0, b = 0)$ if $\bar{x} = 0$.

Blocked and collapsed Gibbs

Blocking means we sample some latent variables from their **joint** conditional, e.g.,

$$\pi_{\text{Gibbs}}(p, b, p_0 \mid p', b', p_0')$$

(p, b) are **blocked**

$$= p(p_0 \mid p, b, \bar{x})\, p(p, b \mid p_0', \bar{x})$$

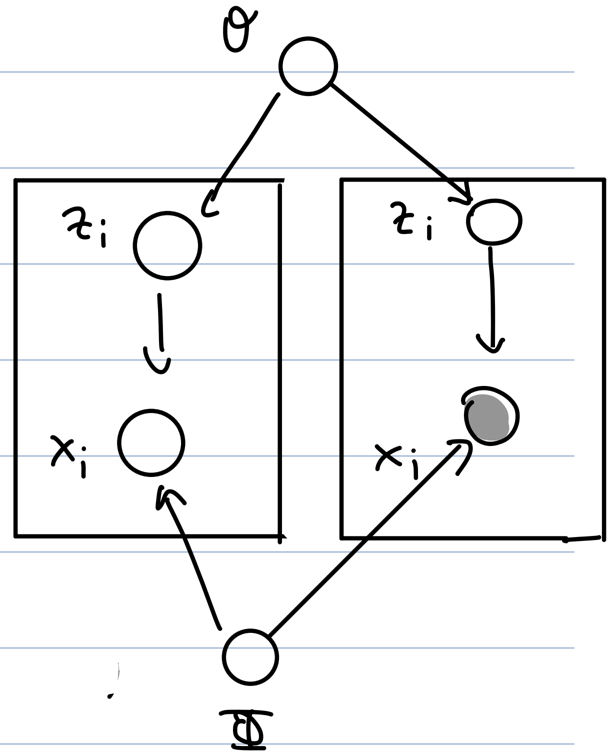
Related to this is **collapsing** where we sample a variable with another variable marginalized out

$$= p(p_0 \mid p, b, \bar{x})\, p(p \mid b, p_0', \bar{x})\, p(b \mid p_0', \bar{x})$$

p is "collapsed" out

This chain is no longer non-ergodic.

Missing data



$$X = X_{obs} \cup X_{miss}$$

Treating x_{miss} as a latent variable

Gibbs w/ missing data

init $z^0, \theta^0, \Phi^0, x^0_{miss}$

for iteration m:

$$z^m \sim P(z \mid \theta^{m-1}, \Phi^{m-1}, x^{m-1}_{miss}, x_{obs})$$

$$\theta^m \sim \dots$$

$$\Phi^m \sim \dots$$

$$x^m_{miss} \sim \underbrace{P(x_{miss} \mid z^m, \theta^m, \Phi^m, x_{obs})}_{\text{Question: how do we sample this?}}$$

Question: $\lim_{m \to \infty} \frac{1}{m} x^m_{miss} = ?$

$$= \mathbb{E}[x_{miss} \mid x_{obs}]$$

(posterior predictive expectation)

Geweke testing (2004)

Consider $X = X_{miss}$ (X_{obs} is empty).

What would be the stationary dist. of the Gibbs sampler above?

$$S \triangleq \{z, \theta, \Phi\}$$

$$\pi^*(S, x_{miss}) = P(S, \underbrace{X_{miss}}_{=X} \mid \underbrace{X_{obs}}_{=\text{empty}}) \equiv P(S, x)$$

the joint!

This suggest a way to test our Gibbs sampler.

Define the <u>forward</u> sampler to be:

for $m = 1 \ldots M$:

$$S_f^m \sim P(S)$$
$$X_f^m \sim P(X \mid S_f^m)$$

And the <u>backward</u> sampler to be Gibbs

for $m = 1 \ldots M$:

$$S_b^m \sim \pi(S_b^m \mid S_b^{m-1}, X_b^{m-1})$$
$$X_b^m \sim P(X \mid S_b^m)$$

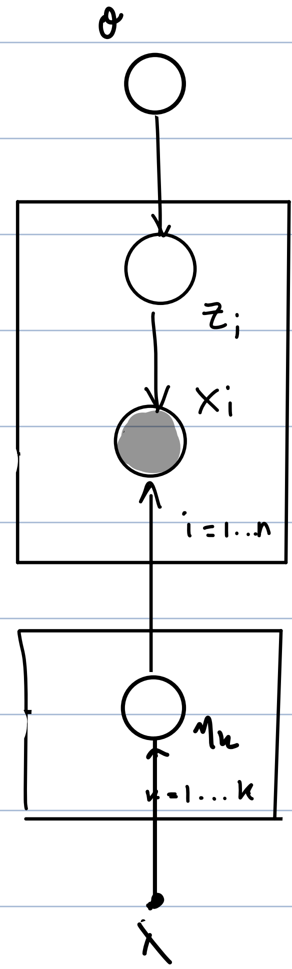
These should both produce samples from the joint:

$$(X_f^m, S_f^m) \sim P(X, S)$$

$$(X_f^b, S_f^b) \sim P(X, S)$$

We can implement both and test whether the samples they generate have the same distribution. If not, there's a bug.

Collapsed Gibbs in the expfam mixture



$p(z_i = k) = \theta_k$

$p(x_i \mid z_i = j) = G(x_i ; \eta_j)$

$F(\eta_k ; \lambda)$

Collapse out $\eta_1 \ldots \eta_K$ when Gibbs sampling $z_1 \ldots z_n$:

$$p(z_i = j \mid x, z_{\setminus i}, \theta)$$

$$\propto \theta_j \, p(x_i \mid z_i = j, z_{\setminus i}, x_{\setminus i})$$

$$\propto \theta_j \int p(x_i, \eta_j \mid z_i = j, z_{\setminus i}, x_{\setminus i}) \, d\eta_j$$

$$\propto \theta_j \int \underbrace{p(x_i \mid \eta_j)} \, \underbrace{p(\eta_j \mid z_{\setminus i}, x_{\setminus i})} \, d\eta_j$$

$$= \underbrace{G(x_i ; \eta_j) \, F(\eta_j ; \lambda_{j,\setminus i}) \, d\eta_j}_{\text{posterior predictive}}$$

$$\lambda_{j,\setminus i} \equiv \begin{bmatrix} \lambda_{j,\setminus i,1} \\ \lambda_{j,\setminus i,2} \end{bmatrix} \begin{matrix} \triangleq \\ \triangleq \end{matrix} \begin{bmatrix} \lambda_1 + \sum_{i' \neq i} t(x_{i'}) f_{i'j} \\ \lambda_2 + \sum_{i' \neq i} f_{i'j} \end{bmatrix}$$

$$\propto \theta_j \frac{\exp\left(a_\ell([\lambda_{j,\setminus i,1} + t(x_i), \lambda_{j,\setminus i,2} + 1])\right)}{\exp\left(a_\ell([\lambda_{j,\setminus i,1}, \lambda_{j,\setminus i,2}])\right)}$$