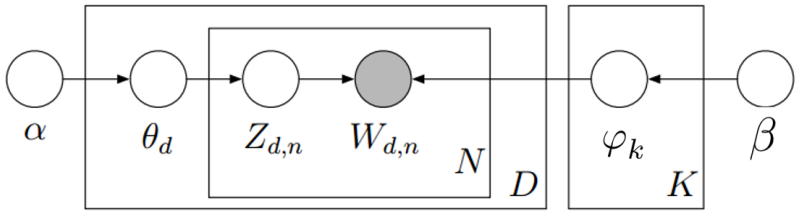
Gibbs Sampler Derivation for Latent Dirichlet Allocation (Blei et al., 2003)

Lecture Notes

Arjun Mukherjee (UH)

I. Generative process, Plates, Notations



1. Draw each topic $\varphi_i \sim \text{Dir}(\eta)$, for $i \in \{1, \ldots, K\}$.
2. For each document:
   1. Draw topic proportions $\theta_d \sim \text{Dir}(\alpha)$.
   2. For each word:
      1. Draw $Z_{d,n} \sim \text{Cat}\ (\theta_d)$.
      2. Draw $W_{d,n} \sim \text{Cat}\ (\varphi_{z_{d,n}})$.

Notations:
D: # of documents
N: # of words/document (N_d for d^{th} document)
K: # of topics
$\theta_d =< \theta_{d,t} | t \in \{1 \ldots K\} >$: topic distribution for document d
$\varphi_k =< \varphi_{k,v} | k \in \{1 \ldots V\} >$: topic k's word distribution over vocabulary V (set of all words)
$z_{d,n}$: latent topic assignment to n^{th} word of document d
$w_{d,n}$: n^{th} word of document d
$Z = \{z_{d,n}\}$, $W = \{w_{d,n}\}$, $\Theta = \{\theta_d\}$, $\Phi = \{\varphi_k\}$

II. The joint distribution:

$$P(W, Z; \alpha; \beta) = P(W|Z)P(Z) = I_1 \times I_2$$

$$I_1 = P(W|Z) = \int (P(W|Z, \Phi)P(\Phi))d\Phi$$

$$I_2 = P(Z) = \int P(Z|\Theta)P(\Theta)d\Theta$$

Let us solve for I_2 first

$$I_2 = P(Z) = \int P(Z|\Theta)P(\Theta)d\Theta$$

From definition, $\Theta = \{\theta_1, \theta_2, \ldots \theta_D\} < \theta_d | d \in \{1 \ldots D\} >$

$$P(\Theta) = \prod_{d=1}^{D} P(\theta_d|\alpha)$$

Now, we know, $\theta_d \sim Dir(\alpha)$. Recall that $\theta_d = \{\theta_{d,1}, \theta_{d,2}, \ldots \theta_{d,K}\} < \theta_{d,k} | t \in \{1 \ldots K\} >$

Recall that $X =< x_1, \ldots x_K > \sim Dir(\alpha_1, \ldots \alpha_K)$ has the following PDF:

$f(x_1, \ldots x_K;\ \alpha_1, \ldots \alpha_K) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} (x_i)^{\alpha_i - 1}$, where $B(\boldsymbol{\alpha}) = \frac{\prod \Gamma(\alpha_i)}{\Gamma(\sum \alpha_i)}$

Also since integrating the PDF over the simplex must equal 1 (from definition), we have

$$\int \left[\frac{1}{B(\alpha)} \prod_{i=1}^{K} (x_i)^{\alpha_i - 1}\right] dX = 1 \quad \text{or} \int \left[\prod_{i=1}^{K} (x_i)^{\alpha_i - 1}\right] dX = B(\boldsymbol{\alpha}) \qquad \text{(Identity A)}$$

Thus,

$P(\Theta) = \prod_{d=1}^{D} P(\theta_d|\alpha) = \prod_{d=1}^{D} \left(\frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} ((\theta_{d,k})^{\alpha-1}\ )\ \right)$ (1)

In our case, we assume each hyper-parameter of the K dimensional Dir, $\alpha_1 = \cdots = \alpha_K = \alpha$,

i.e., the vector $\boldsymbol{\alpha} =< \alpha_1, \ldots \alpha_K > = < \alpha, \ldots \alpha >$

Recall that

$Z = \{z_{d=1,j=1}, \ldots z_{d=1,j=N_d}, \ldots \ldots z_{d=D,j=N_D}\} =< z_{d,j} | d \in \{1 \ldots D\}, j \in \{1 \ldots N_d\}\ >$

$P(Z|\Theta) = \prod_{d=1}^{D} \left(\prod_{j=1}^{N_d} p(z_{d,j}|\theta_d)\right)$ (2)

Now since $z_{d,j} \sim Cat(\theta_d)$

We have $p(z_{d,j}|\theta_d) = \prod_{k=1}^{K} (\theta_{d,k})^{x_k^j}$ where for a given word, j, out of $\{x_{k=1}^j, \ldots x_{k=K}^j\}$ exactly one of $x_k^j = 1$ and rest all are zero, i.e., $x_{\neg k}^j = 0$. This follows directly from the categorical distribution because we are sampling a single topic for the j^{th} word in document d. The sampled topic *k* has the probability of $\theta_{d,k}$.

Thus,

$\prod_{j=1}^{N_d} p(z_{d,j}|\theta_d) = \prod_{j=1}^{N_d} \left(\prod_{k=1}^{K} (\theta_{d,k})^{x_k^j}\right) = \prod_{k=1}^{K} (\theta_{d,k})^{\sum_{j=1}^{N_d} x_k^j}$ (3)

as $y^a \times y^b = y^{a+b}$ where $y = \theta_{d,k}$

Now, $\sum_{j=1}^{N_d} x_k^j$ = # of words in document d that were assigned to topic k. Let us denote this count by $C(d,k)$ *or* C_d^k, i.e.,

$C(d,k) = C_d^k = \sum_{j=1}^{N_d} x_k^j$.

Continuing from (3), we get

$\prod_{j=1}^{N_d} p(z_{d,j}|\theta_d) = \prod_{k=1}^{K} (\theta_{d,k})^{\sum_{j=1}^{N_d} x_k^j} = \prod_{k=1}^{K} (\theta_{d,k})^{C(d,k)}$ (4)

Substituting (4) in (2), we get

$P(Z|\Theta) = \prod_{d=1}^{D} \left(\prod_{j=1}^{N_d} p(z_{d,j}|\theta_d)\right) = \prod_{d=1}^{D} \left(\prod_{k=1}^{K} (\theta_{d,k})^{C(d,k)}\right)$ (5)

Using (5) and (1), we get

$I_2 = P(Z) = \int P(Z|\Theta) P(\Theta) d\Theta$

$= \int \left[\prod_{d=1}^{D} \left(\prod_{k=1}^{K} (\theta_{d,k})^{C(d,k)}\right)\right] \left[\prod_{d=1}^{D} \left(\frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} ((\theta_{d,k})^{\alpha-1}\ )\ \right)\right] d\Theta$ (6)

Since, the document topic distributions, θ_d are independent of each other, we can group as follows:

$= \frac{1}{B(\boldsymbol{\alpha})} \prod_{d=1}^{D} \left[\int \left(\prod_{k=1}^{K} ((\theta_{d,k})^{\alpha-1+C(d,k)}\ )\right) d\theta_d\right]$ (7)

Using identity A, we get

$\int \left(\prod_{k=1}^{K} ((\theta_{d,k})^{\alpha-1+C(d,k)}\ )\right) d\theta_d = B(\boldsymbol{\alpha} + \boldsymbol{C_d})$ where $\boldsymbol{C_d} =< C_d^{k=1}, C_d^{k=2}, \ldots C_d^{k=K} >$

And $B(\boldsymbol{\alpha} + \boldsymbol{C_d}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k + C_d^k)}{\Gamma(\sum_{k=1}^{K} (\alpha_k + C_d^k))}$. Continuing, from (7), we get

$$I_2 = P(Z) = \int P(Z|\Theta)P(\Theta)d\Theta \; = \; \frac{1}{B(\boldsymbol{\alpha})} \prod_{d=1}^{D} B(\boldsymbol{\alpha} + \boldsymbol{C_d}) \quad (8)$$

Now Simplifying I_1

$$I_1 = \; P(W|Z) = \int (P(W|Z, \Phi)P(\Phi))d\Phi$$

$$P(\Phi) = \prod_{k=1}^{K} P(\varphi_k|\beta)$$

Since $\varphi_k \sim Dir(\beta)$

$$P(\Phi) = \prod_{k=1}^{K} \Big(\frac{1}{B(\boldsymbol{\beta})} \prod_{v=1}^{V} ((\varphi_{k,v})^{\beta - 1} \;) \; \Big) \quad (9)$$

$$P(W|Z, \Phi) = \prod_{d=1}^{D} \Big(\prod_{j=1}^{N_d} \big(\prod_{k=1}^{K} p(w_{d,j}|\varphi_k) \big) \Big) \quad (10)$$

Since $w_{d,j} \sim Cat(\varphi_k)$

$p(w_{d,j}|\varphi_k) = \prod_{v=1}^{V} (\varphi_{k,v})^{x_v^{d,j}}$ where for a given word, j, in doc d out of $\{x_{v=1}^{d,j}, \ldots x_{v=V}^{d,j}\}$ exactly one of $x_v^{d,j} = 1$ and rest all are zero, i.e., $x_{\neg v}^{d,j} = 0$ as we are drawing a word v from the topic word distribution, φ_k with probability $\varphi_{k,v}$

Continuing from (10) and substituting $p(w_{d,j}|\varphi_k) = \prod_{v=1}^{V} (\varphi_{k,v})^{x_v^{d,j}}$

$$P(W|Z, \Phi) = \prod_{d=1}^{D} \left(\prod_{j=1}^{N_d} \left(\prod_{k=1}^{K} \prod_{v=1}^{V} (\varphi_{k,v})^{x_v^{d,j}} \right) \right)$$

$$= \prod_{k=1}^{K} \prod_{v=1}^{V} (\varphi_{k,v})^{\sum_d \sum_j x_v^{d,j}}$$

Now, $\sum_d \sum_j x_v^{d,j}$ = # of times word v was assigned to topic k. Let this count be denoted by C_k^v

$$C(k, v) = C_k^v = \sum_d \sum_j x_v^{d,j}$$

Thus, we have

$$P(W|Z, \Phi) = \; \prod_{k=1}^{K} \prod_{v=1}^{V} (\varphi_{k,v})^{C(k,v)} \quad (11)$$

Using (9) and (11), we get

$$I_1 = \; P(W|Z) = \int \big(P(W|Z, \Phi)P(\Phi) \big) d\Phi$$

$$= \; \int \left(\prod_{k=1}^{K} \prod_{v=1}^{V} (\varphi_{k,v})^{C(k,v)} \right) \left(\prod_{k=1}^{K} \left(\frac{1}{B(\boldsymbol{\beta})} \prod_{v=1}^{V} ((\varphi_{k,v})^{\beta - 1} \;) \right) \right) d\Phi$$

Noting that topic distributions, φ_k are independent and grouping the terms

$$= \frac{1}{B(\boldsymbol{\beta})} \prod_{k=1}^{K} \left(\int \prod_{v=1}^{V} (\varphi_{k,v})^{\beta + \; C(k,v) - 1} d\varphi_k \right)$$

The integral simplifies to $B(\boldsymbol{\beta} + \boldsymbol{C_k})$ where $\boldsymbol{C_k} = < C_k^{v=1}, C_k^{v=2}, \ldots C_k^{v=V} >$. Thus,

$I_1 = \ P(W|Z) = \int \big(P(W|Z,\Phi)P(\Phi)\big)d\Phi = \frac{1}{B(\boldsymbol{\beta})}\prod_{k=1}^{K} B(\boldsymbol{\beta} + \boldsymbol{C_k})$ (12)

Thus, using (8) and (12), the full joint distribution of the model can be written as

$P(W, Z; \alpha; \beta) = P(W|Z)P(Z) = I_1 \ \times I_2$

$= \ \Big[\frac{1}{B(\boldsymbol{\alpha})}\prod_{d=1}^{D} B(\boldsymbol{\alpha} + \boldsymbol{C_d})\Big]\Big[\frac{1}{B(\boldsymbol{\beta})}\prod_{k=1}^{K} B(\boldsymbol{\beta} + \boldsymbol{C_k})\Big]$ (13)

III. The Posterior on θ and φ

(A) Computing the posterior distribution for θ_d having *observed topic assignments*, $z_{d,n}$ in document d.

Prior: $\theta_d|\alpha \ \sim Dir\ (\alpha)$. Prior probability: $P(\theta_d) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{k=1}^{K}((\theta_{d,k})^{\alpha-1}\ )$

Likelihood: $z_{d,j}|\theta_d \ \sim Cat(\theta_d)$.

Likelihood of this document given θ_d: $P(Z_d|\theta_d)\prod_{j=1}^{N_d} P(z_{d,j}|\theta_d) = \prod_{k=1}^{K}(\theta_{d,k})^{C(d,k)}$ (using (4))

Posterior on θ_d using Bayes theorem:

$$P(\theta_d|Z_d) = \frac{P(Z_d|\theta_d)P(\theta_d)}{\int(P(Z_d|\theta_d)P(\theta_d))d\theta_d} \propto \prod_{k=1}^{K}(\theta_{d,k})^{C(d,k)+\alpha-1}$$

i.e., Posterior on $\theta_d|Z_d \sim Dir(\boldsymbol{\alpha} + \boldsymbol{C_d})$

which is nothing but proportional to the $Dirichlet(\boldsymbol{\alpha} + \boldsymbol{C_d})$. Here we see conjugate priors in action. Since Dirichlet is the conjugate prior of Categorical distribution, the posterior also takes the form of the prior, i.e., another Dirichlet with added pseudocounts.

In fact one can directly use the properties of conjugate priors to arrive at $\theta_d|Z_d \sim Dir(\boldsymbol{\alpha} + \boldsymbol{C_d})$.

Using the properties of the Dirichlet distribution, one can easily obtain the expected value of the probability mass associated to each topic in the document d as follows:

$$E[\theta_{d,k}] = \ \widehat{\theta_{d,k}} = \ \frac{\alpha + C(d,k)}{\sum_{k=1}^{K}(\alpha + C(d,k))}$$ (14)

(B) Computing the posterior distribution for φ_k

Using the fact that the Dirichlet is conjugate to the categorical distribution used for word emission process, we can arrive at $\varphi_k|W_k \sim Dir(\boldsymbol{\beta} + \boldsymbol{C_k})$. Thus,

$$E[\varphi_{k,v}] = \ \widehat{\varphi_{k,v}} = \ \frac{\beta + C(k,v)}{\sum_{v=1}^{V}(\beta + C(k,v))}$$ (15)

IV. Gibbs sampler

Let $Z \ = \ \{z_{d,n}\} = \{z_{d=1,j=1}, \ldots z_{d=1,j=N_1}, \ldots z_{d=D,j=N_D}\}$ be a $\sum N_i$ dimensional random vector, i.e., Z denotes the collection of all latent topic variables, $z_{d,n}$ corresponding to all words in all documents.

Also let us posit a Markov chain $X = < Z^{(0)}, Z^{(1)}, Z^{(2)}, \ldots . Z^{(N_{Iter})} >$ over the data and the model, whose stationary distribution converges to the posterior on distribution of Z.

Also let $Z = \{z_{d,n}\} = \{z_i\}$ be denoted by single subscript of ease of notation. For a given token/word, $i = (d', j')$, i.e., the word j' at document d', the Gibbs conditional (sampling distribution) for its latent topic z_i can be constructed as follows:

$$P(z_i = k'|Z_{\neg i}, W) = \frac{P(Z,W)}{P(Z_{\neg i},W)} = \frac{P(W|Z)P(Z)}{P(W_{\neg i}|Z_{\neg i})P(Z_{\neg i})p(w_i)} \propto \frac{P(W|Z)P(Z)}{P(W_{\neg i}|Z_{\neg i})P(Z_{\neg i})} \quad (16)$$

where $Z = \{z_i, Z_{\neg i}\}$ and $W = \{w_i, W_{\neg i}\}$. Equation (16) gives us the probability that the latent topic variable z_i at $i = (d', j')$ is assigned to topic $k' \in \{1 \dots K\}$ having observed all other topic assignments and words except w_i. Also for subsequent steps the subscript $\neg i$ denotes all counts/variables/functional values upon discounting (or not accounting) the token at w_i

Expanding the sampling distribution using (13) we get:

$$P\left(z_i = k^{'} \middle| Z_{\neg i}, W_{\neg i}, w_i\right) \propto \frac{P(W|Z)P(Z)}{P(W_{\neg i}|Z_{\neg i})P(Z_{\neg i})} \propto \frac{\left[\left(\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})\right)\left(\prod_{k=1}^{K} B(\boldsymbol{\beta}+\boldsymbol{C_k})\right)\right]}{\left[\left(\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})\right)\left(\prod_{k=1}^{K} B(\boldsymbol{\beta}+\boldsymbol{C_k})\right)\right]_{\neg i}}$$

$$\propto \left[\frac{\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})}{\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})_{\neg i}}\right] \times \left[\frac{\prod_{k=1}^{K} B(\boldsymbol{\beta}+\boldsymbol{C_k})}{\prod_{k=1}^{K} B(\boldsymbol{\beta}+\boldsymbol{C_k})_{\neg i}}\right] \quad (17)$$

Thus, the Gibbs sampling distribution for $p(z_i = k)$ says that it is proportional to the full joint distribution of the model divided by the joint considering the token/word, w_i and its associated topical assignment did not exist in our data/model.

Observing that α remains fixed, and i corresponds to the topic and words, $\{z_i, w_i = (d', j')\}$ at some document d' and some position j' in that document, we can further simplify the first term in (17) as

$$\left[\frac{\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})}{\prod_{d=1}^{D} B(\boldsymbol{\alpha}+\boldsymbol{C_d})_{\neg i}}\right] = \left[\frac{B(\boldsymbol{\alpha}+\boldsymbol{C_{d=d'}})}{\left(B(\boldsymbol{\alpha}+\boldsymbol{C_{d=d'}})\right)_{\neg i}}\right] = \frac{B(\boldsymbol{\alpha}+\boldsymbol{C_{d=d'}})}{B(\boldsymbol{\alpha}+[\boldsymbol{C_{d=d'}}]_{\neg i})} \quad (18)$$

As for all documents other than d' the numerator and denominator remain exactly the same and cancel out. Now let us see what changes happen in the count vector $\boldsymbol{C_{d=d'}}$ with or without including the term/word at $w_i = (d', j')$ whose latent topic assignment is $z_i = k'$. We recall that $\boldsymbol{C_d} =< C_d^{k=1}, C_d^{k=2}, \dots C_d^{k=K} >$ and $C_d^k = C(d,k)$ # of words in document d that were assigned to topic k. Hence, we can write

$$C(d,k) = \begin{cases} [C(d,k)]_{\neg i} & ; k \neq k' \\ [C(d,k)]_{\neg i} + 1 & ; k = k' \end{cases} \quad (19)$$

Expanding (18) using the definition of $B(\boldsymbol{\alpha}) = \frac{\prod \Gamma(\alpha_i)}{\Gamma(\sum \alpha_i)}$, we get

$$\frac{B(\boldsymbol{\alpha}+\boldsymbol{C_{d=d'}})}{B(\boldsymbol{\alpha}+[\boldsymbol{C_{d=d'}}]_{\neg i})}$$

$$= \frac{\prod_{k=1}^{K} \Gamma(\alpha+C(d',k))}{\prod_{k=1}^{K} \Gamma(\alpha+C(d',k)_{\neg i})} \times \frac{\Gamma\left(\sum_{k=1}^{K}(\alpha+C(d',k)_{\neg i})\right)}{\Gamma\left(\sum_{k=1}^{K}(\alpha+C(d',k))\right)}$$

Using the result in (19) this further simplifies to (upon canceling out all non k' terms)

$$= \frac{\Gamma(\alpha + C(d',k'))}{\Gamma(\alpha + C(d',k')_{\neg i})} \times \frac{\Gamma\left(\sum_{k=1\backslash k'}^{K}(\alpha + C(d',k)_{\neg i}) + \ [\alpha + C(d',k')_{\neg i}]\right)}{\Gamma\left(\sum_{k=1\backslash k'}^{K}(\alpha + C(d',k)) + \ [\alpha + C(d',k')]\right)}$$

$$= \frac{\Gamma(\alpha + C(d',k')_{\neg i} + 1)}{\Gamma(\alpha + C(d',k')_{\neg i})} \times \frac{\Gamma\left(\sum_{k=1\backslash k'}^{K}(\alpha + C(d',k)_{\neg i}) + \ [\alpha + C(d',k')_{\neg i}]\right)}{\Gamma\left(\sum_{k=1\backslash k'}^{K}(\alpha + C(d',k)_{\neg i}) + \ [\alpha + C(d',k')_{\neg i} + 1]\right)}$$

Using the identity $\Gamma(x+1) = x\Gamma(x)$, we get

$$\frac{B(\boldsymbol{\alpha} + \boldsymbol{C}_{\boldsymbol{d}=\boldsymbol{d}'})}{B(\boldsymbol{\alpha} + [\boldsymbol{C}_{\boldsymbol{d}=\boldsymbol{d}'}]_{\neg i})} = \frac{\alpha + C(d', k')_{\neg i}}{\sum_{k=1}^{K}(\alpha + C(d', k)_{\neg i})}$$

In a similar way, one can also simply the second term in (17) as follows:

$$\left[\frac{\prod_{k=1}^{K} B(\boldsymbol{\beta} + \boldsymbol{C}_{\boldsymbol{k}})}{\prod_{k=1}^{K} B(\boldsymbol{\beta} + \boldsymbol{C}_{\boldsymbol{k}})_{\neg i}}\right] = \frac{B\left(\boldsymbol{\alpha} + \boldsymbol{C}_{\boldsymbol{k}=\boldsymbol{k}'}\right)}{\left(B\left(\boldsymbol{\alpha} + \boldsymbol{C}_{\boldsymbol{k}=\boldsymbol{k}'}\right)\right)_{\neg i}} = \frac{\beta + C\left(k', v'\right)_{\neg i}}{\sum_{v=1}^{V}\left(\beta + C\left(k', v'\right)_{\neg i}\right)}$$

Where v' refers to the vocabulary token which is assigned to w_i. We can thus simplify the full Gibbs conditionals as follows:

$$P(z_i = k' | Z_{\neg i}, W_{\neg i}, w_i) \propto \left[\frac{\alpha + C(d', k')_{\neg i}}{\sum_{k=1}^{K}(\alpha + C(d', k)_{\neg i})}\right] \times \left[\frac{\beta + C(k', v')_{\neg i}}{\sum_{v=1}^{V}(\beta + C(k', v')_{\neg i})}\right]$$

The full algorithm for Gibbs sampling is as follows (from Fig 8, [Heinrich, 2008]). There are some minor notation changes which are noted below.

$$D \to M; C_d^k \to n_m^{(k)}; C_k^v \to n_k^{(t)}; n_m = \sum_k n_m^{(k)}\ ;\ n_m = \sum_t n_k^{(t)}$$

⊡ initialisation
zero all count variables, $n_m^{(k)}, n_m, n_k^{(t)}, n_k$
for all documents $m \in [1, M]$ **do**
  for all words $n \in [1, N_m]$ in document m **do**
    sample topic index $z_{m,n}$=$k \sim$ Mult(1/K)
    increment document–topic count: $n_m^{(k)} + 1$
    increment document–topic sum: $n_m + 1$
    increment topic–term count: $n_k^{(t)} + 1$
    increment topic–term sum: $n_k + 1$
  end for
end for
⊡ Gibbs sampling over burn-in period and sampling period
while not finished **do**
  for all documents $m \in [1, M]$ **do**
    for all words $n \in [1, N_m]$ in document m **do**
      ⊡ for the current assignment of k to a term t for word $w_{m,n}$:
      decrement counts and sums: $n_m^{(k)} - 1; n_m - 1; n_k^{(t)} - 1; n_k - 1$
      ⊡ multinomial sampling acc. to Eq. 79 (decrements from previous step):
      sample topic index $\tilde{k} \sim p(z_i | \vec{z}_{\neg i}, \vec{w})$
      ⊡ use the new assignment of $z_{m,n}$ to the term t for word $w_{m,n}$ to:
      increment counts and sums: $n_m^{(\tilde{k})} + 1; n_m + 1; n_{\tilde{k}}^{(t)} + 1; n_{\tilde{k}} + 1$
    end for
  end for
  ⊡ check convergence and read out parameters
  if converged and L sampling iterations since last read out **then**
    ⊡ the different parameters read outs are averaged.
    read out parameter set $\underline{\Phi}$ according to Eq. 82
    read out parameter set $\underline{\Theta}$ according to Eq. 83
  end if
end while
