HW5B

1) basic sampling (unif, rejection, inverse transform ...)

2) Gibbs sampling

3) Dirichlet-sampling (Gibbs)

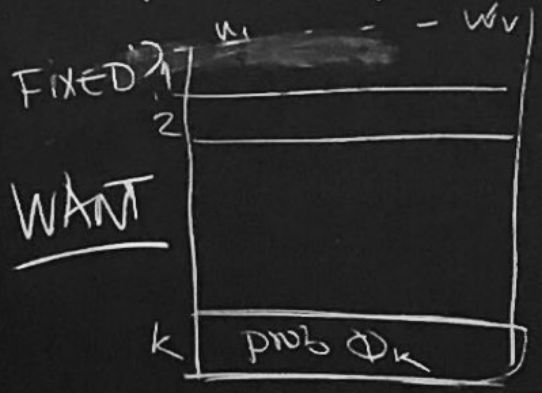
4) baby LDA inference, given data ⟹ find params

- 20NG V = vocab-size K = 13,000
- by dataset 1000 docs, 500 unique words (debug) K = 20

LDA as generative process

Given params $\Rightarrow$ generate data (fake)

k = #topics



topic $\phi^k = \phi_k$ = distribution (categorical) $(\text{words}^V) = (\phi_{k1}^1, \phi_{k2}^2, \ldots \phi_k^V)$ probs over words

NOT FIXED
topics = unknown?

ϕ^k = R.V. topic $\sim$ distributed given prior Dirichlet(β)

$(\beta^1, \ldots \beta^V)$ β = prior params over V words

$d \in D$ documents θ_d = distribution over topics $= (\theta_d^1, \theta_d^2, \ldots \theta_d^K)$ probs over topics $\sum_{j=1}^{k} \theta_d^j = 1$

unknown

θ_d = R.V. dist $\sim$ distributed given prior Dirichlet(α)

α = prior params

$(\alpha^1, \alpha^2, \ldots \alpha^k)$ over k topics

Generate data

- Sample k topics $\phi^k \sim$ Dirichlet(β)

For each doc $d \in D$

- Sample $\theta_d \sim$ Dirichlet(α) θ_d = dist(topics)
- L_d = (given) known length of doc d
- For $i = 1 : L_d$ i = index of word i-th word w_i
  - Sample topic for word w_i $z_i \sim$ Multinomial (θ_d) (categorical?)
  - Sample word w from topic z_i: $w_i \sim$ Multinomial (ϕ_{z_i})

categ dist (discrete)

probs $p_1, p_2 \ldots p_V$

$\sum p_i = 1$

$\text{prob}(X = k) = p_k$

$V=2$ BINOMIAL (WIN)

N trials for V categories

output is $n_1 + n_2 + \ldots n_V = N$

$$P(n_1, n_2 \ldots n_V \mid N, p_1, p_2 \ldots p_V) = \frac{N!}{n_1! n_2! \ldots n_V!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdots p_V^{n_V}$$

COIN BIN (H 0.3, T 0.7)

N = 20 trials

MULTINOM

INFERENCE

- start with initial $\bar{\alpha} = (\alpha_1 \ldots \alpha_K)$ $\bar{\beta} = (\beta_1, \ldots \beta_V)$
- generative process (implicit) $\rightarrow$ $\bar{d}_j$ = bag-of-words-docs $d_j = (d_j^1, d_j^2, \ldots d_j^V)$ $\}$ $j = 1..D$

  $d_j - \Theta_j$ = dist over topics

- Likelihood $P(D = \text{docs} \mid \alpha, \beta, K, L_{avg}) = OBJ$ (params; K, L_{avg} fixed)

WANT: $\arg\max_{(\alpha, \beta)} P(D, \alpha, \ldots)$

α = prop with dist $\alpha(1, 1, 2, 3)$ $\left(\frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}\right)$

V=2 COIN = R.V (p, 1-p) p-?

(prior) (initial guess)
(belief w/out ...)

of probab

distrib over possible (p, 1-p)

coin bias	50p - 50(1-p)	60 — 40	70 — 30	90 — 10
prob of (p, 1-p)	60%	10%	10%	20%

$$\text{Beta}\left(\begin{matrix}p\\1-p\end{matrix}\ \middle|\ \begin{matrix}\text{params}\\a, b\end{matrix}\right) = \frac{p^{a-1}\cdot(1-p)^{b-1}\cdot\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}$$

bias | params

$\Gamma(a) = (a-1)!$

$\Gamma(b) = (b-1)!$

$\Gamma(a+b) = (a+b-1)!$

$\Gamma(x+1) = \text{"}(x+1)!\text{"} = x! \cdot (x+1)$

$= \Gamma(x)\,(x+1)$

a=2, b=1	n_1=6H	n_2=4T	a'=8, b'=5
a=2, b=1	n_1=500	n_2=500	a'=502, b'=501
a=2000, b=1000	500 H	500 T	a'=2500, b'=1500

$p(N = n_1 + n_2 \mid \text{prior-beta } a, b) \simeq$ Multi-Dirichlet

$\text{prob}\left(\begin{matrix}p\\1-p\end{matrix}\ \middle|\ \text{prior } a, b,\ n_1, n_2\right) \propto \text{Beta}(a+n_1, b+n_2)$