HW #4

1) Query likelihood model:

$$score(Q, D) = \sum logp(q_i|d)$$

$$= \sum c(w_i, q) logp(w_i|\theta_D)$$

We know that : $c(w_i, Q) = |Q|p(w|\theta_Q)$

$$score(Q, D) = \sum |Q|p(w|\theta_Q) logp(w_i|\theta_D)$$

$|Q|$ is a constant for a given query

$$score(Q, D) = \sum p(w|\theta_Q) logp(w_i|\theta_D)$$

$$= -D(\theta_Q||\theta_D)$$

$\therefore$ KL divergence covers query likelihood estimate as a special case when we set $p(w|\theta_Q) = \frac{c(w_i,Q)}{|Q|}$

2) a) Multinomial distribution

$$p(q|d) \prod p(w_i = x_i|d_)$$

Since the probability of each word is multiplied together, both w_1 and w_2 must occur. $\therefore$ $\boxed{\text{conjuctive}}$

b)

$$p(Q|d) = \prod_{i=1}^{k} p(q_i = Q_i|d)$$

$$= \prod_{i=1}^{k} \sum_{j=1}^{n_i} p(w_i = x_i|d)$$

3)a)

$$p(\text{“}the\text{”}) = p(w = \text{“}the\text{”}|H) + p(w = \text{“}the\text{”}|T)$$

$$= 0.8 * 0.3 + 0.3 * 0.2 = \boxed{0.3}$$

b) It doesn't matter whether its the 2nd word or the 1st. Since we do not have constraint on what the 1st word must be, p=1 for whatever happens to the first word. ∴ $\boxed{\text{p(w\_2="the") = 0.3}}$

c)

$$p(H|w = \text{"}data\text{"}) = \frac{p(w = \text{"}data\text{"}|H)}{p(w = \text{"}data\text{"})}$$

$$= \frac{p(w = \text{"}data\text{"}|H)}{p(w = \text{"}data\text{"}|H) + p(w = \text{"}data\text{"}|T)}$$

$$= \frac{0.1 * 0.8}{0.1 * 0.8 + 0.1 * 0.2} = \boxed{0.8}$$

d) Compare p(w):

$$p(w) = p(w|H) + p(w|T)$$

We expect that p(w="data") is lowest since both p(w|H), p(w|T) is lowest among the 5 words. So the word "data" is expected to occur least frequently.

e) From the data, we observe: p("the") = 3/10
p("computer") = 3/10
p("data") = 2/10
p("game") = 2/10 We also know that the probability of flipping 10 heads in a row is $p = 10C10 * 0.8^{10} * 0.2^0 = 0.8^10$. The estimated probability of p("computer"|H) = 3/10*0.8 =0.24 and p("game"|H) = 2/10*0.8 =0.16.

4) a) $p(w) = (1 - \lambda)p(w|C) + \lambda p(w|\theta_1)$
b)

$$L = \sum c(w, D_1) logp(w)$$

$$= \sum c(w, D_1) log\left[(1 - \lambda)p(w|C) + \lambda p(w|\theta_1)\right]$$

c) We need the hidden variable z_w which determines whether a word w is from the collection C or from the topic θ_1. We use $p(w|\theta_1)$ to estimate $p(z_w = 1)$ and vice versa.

d) E step:

$$p(z_w = 1) = \frac{\lambda p^{(n)}(w|\theta_1)}{\lambda p^{(n)}(w|\theta_1) + (1 - \lambda)p(w|C)}$$

$p(w|\theta_1) = \frac{c(w, D_1)}{|D_1|}$ M-step:

$$p^{(n+1)}(w|\theta_1) = \frac{c(w, D_1)p(z_w = 1)}{\sum_{w \in V} c(w, D_1)p(z_{w'} = 1)}$$

Estimate λ by :

$$\hat{\lambda} = argmax[p(z_w = 1)]$$