1 point

1. (True/False) Each iteration of Gibbs sampling for Bayesian inference in topic models is guaranteed to yield a higher joint model probability than the previous sample.

- ( ) True
- (•) False

---

1 point

2. (Check all that are true) Bayesian methods such as Gibbs sampling can be advantageous because they

- [x] Account for uncertainty over parameters when making predictions
- [ ] Are faster than methods such as EM
- [ ] Maximize the log probability of the data under the model
- [x] Regularize parameter estimates to avoid extreme values

---

1 point

3. For the standard LDA model discussed in the lectures, how many parameters are required to represent the distributions defining the topics?

- ( ) [# unique words]
- (•) [# unique words] * [# topics]
- ( ) [# documents] * [# unique words]
- ( ) [# documents] * [# topics]

---

2 points

4. Suppose we have a collection of documents, and we are focusing our analysis to the use of the following 10 words. We ran several iterations of collapsed Gibbs sampling for an LDA model with K=2 topics and alpha=10.0 and gamma=0.1 (with notation as in the collapsed Gibbs sampling lecture). The corpus-wide assignments at our most recent collapsed Gibbs iteration are summarized in the following table of counts:

Word	Count in topic 1	Count in topic 2
baseball	52	0
homerun	15	0
ticket	9	2
price	9	25
manager	20	37
owner	17	32
company	1	23
stock	0	75
bankrupt	0	19
taxes	0	29

We also have a single document i with the following topic assignments for each word:

topic	1	2	1	2	1
word	baseball	manager	ticket	price	owner

Suppose we want to re-compute the topic assignment for the word "manager". To sample a new topic, we need to compute several terms to determine how much the document likes each topic, and how much each topic likes the word "manager". The following questions will all relate to this situation.

First, using the notation in the slides, what is the value of $m_{\text{manager},1}$ (i.e., the number of times the word "manager" has been assigned to topic 1)?

20

---

1 point

5. Consider the situation described in Question 4.

What is the value of $\sum_w m_{w,1}$, where the sum is taken over all words in the vocabulary?

123

---

1 point

6. Consider the situation described in Question 4.

Following the notation in the slides, what is the value of $n_{i,1}$ for this document i (i.e., the number of words in document i assigned to topic 1)?

3

---

1 point

7. In the situation described in Question 4, "manager" was assigned to topic 2. When we remove that assignment prior to sampling, we need to decrement the associated counts.

After decrementing, what is the value of $n_{i,2}$?

1

---

1 point

8. In the situation described in Question 4, "manager" was assigned to topic 2. When we remove that assignment prior to sampling, we need to decrement the associated counts.

After decrementing, what is the value of $m_{manager,2}$?

36

---

1 point

9. In the situation described in Question 4, "manager" was assigned to topic 2. When we remove that assignment prior to sampling, we need to decrement the associated counts.

After decrementing, what is the value of $\sum_w m_{w,2}$?

241

---

2 points

10. Consider the situation described in Question 4.

As discussed in the slides, the unnormalized probability of assigning to topic 1 is

$$p_1 = \frac{n_{i,1}+\alpha}{N_i-1+K\alpha} \frac{m_{\text{manager},1}+\gamma}{\sum_w m_{w,1}+V\gamma}$$

where V is the total size of the vocabulary.

Similarly the unnormalized probability of assigning to topic 2 is

$$p_2 = \frac{n_{i,2}+\alpha}{N_i-1+K\alpha} \frac{m_{\text{manager},2}+\gamma}{\sum_w m_{w,2}+V\gamma}$$

Using the above equations and the results computed in previous questions, compute the probability of assigning the word "manager" to topic 1.

(*Reminder: Normalize across the two topic options so that the probabilities of all possible assignments---topic 1 and topic 2---sum to 1.*)

Round your answer to 3 decimal places.

0.5622