1. What are **Monte Carlo** methods? Briefly describe a Monte Carlo method for computing $\pi$.

2. What is a **Markov chain**? What does it mean for a Markov chain to have a **stationary distribution**?

3. What is a **Markov chain Monte Carlo** (MCMC) algorithm? Name two different MCMC algorithms, and briefly describe the difference between them.

4. The Dirichlet distribution is defined as

$$p(\theta \mid \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} \theta_{i}^{\alpha_{i} - 1}$$

where $B(\alpha)$ is the multivariate beta function. Show that the Dirichlet distribution is the conjugate prior of the multinomial.

5. Prove the following:

$$\int \prod_{i=1}^{K} \theta_{i}^{n_{i} + \alpha_{i} - 1} d\theta = B(\alpha + n)$$
6. Using the fact that the Dirichlet is conjugate to the multinomial, perform the following integration to “collapse” out the prior over $\Theta$ in LDA:

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

7. Now, collapse the prior over $\Phi$ in a similar way:

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

8. Draw the collapsed model using plate notation, and write down its joint distribution as an equation.
9. Explain why the definition of the joint distribution above is sufficient for computing the sampling probability $P(z_{d,i} = k \mid Z_{-d,i}, W, \alpha, \beta)$ needed at each step of the Gibbs sampler.

10. Write down the sampling probability $P(z_{d,i} = k \mid Z_{-d,i}, W, \alpha, \beta)$ used in each step of the collapsed Gibbs sampler. Can you intuitively describe the two parts of the equation?

11. What must be stored in memory to run the collapsed Gibbs sampler?

12. Describe how to infer the marginalized variables $\Theta$ and $\Phi$ from the current state of the Markov chain.