Quiz: Directed Graphical Models & Inference

STA414/2104 - Winter 2026

1. **HMMs with Identity Transition**
   In the Forward Algorithm for Hidden Markov Models (HMMs), suppose the transition matrix A is the identity matrix ($A = I$), such that $A_{ii} = 1$ and $A_{ij} = 0$ for $i \neq j$.

   Which of the following statements best describes the behavior of the filtered marginal $\alpha_t(j) = p(z_t = j|x_{1:t})$?

   (a) It depends only on the current observation x_t, ignoring all past observations.
   (b) It is proportional to the product of the initial probability $\pi(j)$ and the emission probabilities for that state across all time steps up to t.
   (c) It remains constant over time, regardless of the observations $x_{1:t}$.
   (d) It becomes a uniform distribution because the state cannot change.

   Correct Answer: (b)
   Rationale: The Forward Algorithm updates the filtered marginal recursively:

   $$\alpha_t(j) \propto p(x_t|z_t = j) \sum_i p(z_t = j|z_{t-1} = i)\alpha_{t-1}(i)$$

   With $A = I$, the transition term $p(z_t = j|z_{t-1} = i)$ is 1 if $i = j$ and 0 otherwise. This simplifies the recursion to:

   $$\alpha_t(j) \propto p(x_t|z_t = j)\alpha_{t-1}(j)$$

   Unrolling this recursion back to $t = 1$, we get:

   $$\alpha_t(j) \propto p(z_1 = j) \prod_{k=1}^{t} p(x_k|z_k = j)$$

   Thus, the belief state for j is proportional to the initial prior times the likelihood of generating the entire observed sequence from that single, constant state j.

2. **Rejection Sampling Acceptance Probability**
Recall from the lecture that the probability of accepting a specific candidate sample x (where $x \sim q(x)$ and $u|x \sim \text{Unif}[0, cq^*(x)]$) is given by:

$$\mathbb{P}(u \leq p^*(x)|x) = \frac{p^*(x)}{cq^*(x)}$$

where $p(x) = \frac{p^*(x)}{Z_p}$ and $q(x) = \frac{q^*(x)}{Z_q}$.

Given this formula, what is the probability of acceptance, $\mathbb{P}(\text{accept proposal } x)$?

(a) $\frac{p^*(x)}{cq^*(x)}$

(b) $\frac{1}{c}$

(c) $\frac{Z_p}{cZ_q}$

(d) $c\frac{Z_p}{Z_q}$

Correct Answer: (c)
Rationale: The marginal probability of acceptance is found by integrating the conditional acceptance probability with respect to the proposal distribution $q(x)$:

$$\mathbb{P}(accept) = \int \mathbb{P}(accept|x)q(x)dx = \int \frac{p^*(x)}{cq^*(x)}q(x)dx$$

Substituting the definitions $p^*(x) = Z_p p(x)$ and $q^*(x) = Z_q q(x)$:

$$\mathbb{P}(accept) = \int \frac{Z_p p(x)}{cZ_q q(x)}q(x)dx = \frac{Z_p}{cZ_q}\int p(x)dx$$

Since $\int p(x)dx = 1$, the final acceptance probability is $\frac{Z_p}{cZ_q}$.

3. **Quantitative Curse of Dimensionality**
Consider a Rejection Sampling setup where the target $p(\mathbf{x})$ and proposal $q(\mathbf{x})$ factorize over D independent dimensions (i.e., $p(\mathbf{x}) = \prod_{i=1}^{D} p(x_i)$) (with the same marginal distribution for all x_i). Suppose that for a single dimension ($D = 1$), the optimal constant c results in an acceptance rate of 50% (0.5).

If you use this same sampling strategy in $D = 10$ dimensions, what is the expected acceptance rate?

(a) It remains 50% because the dimensions are independent.
(b) 5%, decreasing linearly with dimension (0.5×10^{-1}).
(c) $\approx 0.1\%$, because the acceptance rate scales as 0.5^{10}.
(d) 0%, because rejection sampling is theoretically impossible in dimensions $D > 3$.

Correct Answer: (c)
Rationale: The acceptance rate is given by $\frac{Z_p}{cZ_q}$. In this independent setup, the total acceptance probability is the product of the acceptance probabilities for each dimension. Alternatively, the constant c required to bound the joint distribution is the product of the constants for each dimension: $c_{total} = (c_1)^D$. Since the acceptance rate in 1D is $1/c_1 = 0.5$, the acceptance rate in D dimensions is $(1/c_1)^D = 0.5^{10} \approx \frac{1}{1024} \approx 0.1\%$. This illustrates why acceptance rates are "exponentially small in dimension".

4. **Efficiency of Rejection Sampling**
You are designing a Rejection Sampling algorithm to sample from a target distribution $p(x)$ using a proposal distribution $q(x)$. Recall that you must find a constant c such that $cq(x) \geq p(x)$ for all x.

Consider two different proposal distributions:

- **Proposal A:** $q_A(x)$ is a "broad" distribution (e.g., a Gaussian with very large variance) that covers $p(x)$ but is much flatter.
- **Proposal B:** $q_B(x)$ is a "spiky" distribution (e.g., a Gaussian with very small variance) that matches the peak of $p(x)$ well but decays much faster than $p(x)$ in the tails.

Which of the following outcomes is most likely?

(a) **Proposal A** is valid but inefficient because the required constant c will be very large, resulting in a high rejection rate.

(b) **Proposal B** is the best choice because it fits the high-probability regions of $p(x)$ tightly, minimizing the rejection rate.

(c) **Proposal B** is valid, but it will result in biased samples because it ignores the tails of $p(x)$.

(d) **Proposal A** is invalid because a flat distribution cannot upper-bound a peaked distribution.

Correct Answer: (a)
Rationale: Rejection sampling requires $cq(x) \geq p(x)$ everywhere.

- *For **Proposal B** (tight/light-tailed), if $q(x)$ decays faster than $p(x)$ in the tails, the ratio $p(x)/q(x)$ goes to infinity as $|x| \to \infty$. This makes it impossible to find a finite constant c that satisfies the inequality everywhere, rendering Proposal B **invalid**.*
- *For **Proposal A** (broad/heavy-tailed), the inequality can be satisfied. However, because $q_A(x)$ is much flatter than $p(x)$, the gap between $cq_A(x)$ and $p(x)$ will be massive in most regions. Since the acceptance rate is $1/c$ (assuming normalized densities) or Z_p/cZ_q, a large c leads to extremely rare acceptance.*

Thus, Proposal A is valid but computationally wasteful.