Problem 1: Distributions closed under convolution

1. Let $p = \mathcal{N}(\mu_1, \sigma_1^2)$ and $q = \mathcal{N}(\mu_2, \sigma_2^2)$. Define the convolution as:

$$(p * q)(x) = \int_{-\infty}^{\infty} p(x-y)q(y)dy$$

   Show that $(p * q)(x) = \mathcal{N}(x; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

2. Let $q(x_k \mid x_{k-1}) = \mathcal{N}\left(x_k; \alpha_k x_{k-1}, \sigma_k^2 I\right)$ for $k = 1, \dots, n$. Starting from a fixed data point x_0, show that the conditional distribution $q(x_k \mid x_0)$ is a Gaussian of the form $\mathcal{N}(x_k; \tilde{\mu}_k, \tilde{\sigma}_k^2 I)$. Provide the explicit expressions for $\tilde{\mu}_k$ and $\tilde{\sigma}_k^2$ in terms of α_i, σ_i^2, and x_0.

3. Let $p = \text{Cauchy}(\mu_1, \gamma_1)$ and $q = \text{Cauchy}(\mu_2, \gamma_2)$, where the density is given by:

$$p(x; \mu, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-\mu}{\gamma}\right)^2\right)}$$

   Show that $p * q = \text{Cauchy}(\mu_1 + \mu_2, \gamma_1 + \gamma_2)$, and we can therefore use Cauchy distributions instead of Gaussians in the definition of the noising process.

   Note: for recent work on diffusion models with heavy-tailed distributions, refer to [1] and [2].

Problem 2: Closed-form for ELBO

As shown in the lecture, the ELBO for hierarchical VAE can be optimised using the following objective:

$$\mathcal{L}(\theta) = \mathbb{E}_{z_0 \sim p(z_0), z_{1,\dots,N} \sim q(z_{1,\dots,N} \mid z_0)}\left[\log p(z_N) \prod_{n=1}^{N} p_\theta(z_{n-1} \mid z_n)\right]$$

1. Assuming $p_\theta(z_{n-1} \mid z_n) = \mathcal{N}(z_n; f_\theta(z_n, n), \sigma_{n-1}^2 I)$ and $p(z_N)$ does not depend on θ, show that maximising $\mathcal{L}(\theta)$ is equivalent to minimising the following objective:

$$\widetilde{\mathcal{L}}(\theta) = \mathbb{E}_{\substack{n \sim \text{Unif}(\{1,\dots,N\}), \\ z_0 \sim p(z_0), \\ z_n, z_{n-1} \sim q(z_n, z_{n-1} \mid z_0)}}\left[\frac{1}{\sigma_{n-1}^2}\|z_{n-1} - f_\theta(z_n,\ n)\|^2\right]$$

2. How can we sample from a model trained with the objective $\widetilde{\mathcal{L}}(\theta)$? Write the response as a probabilistic program.

3. What are the possible disadvantages of using the objective $\widetilde{\mathcal{L}}(\theta)$?

Problem 3: Optimal denoiser

Consider the following training objective:

$$\mathcal{L}(f) = \mathbb{E}_{x \sim p(x), z \sim q(z \mid x)}\left[\|x - f(z)\|^2\right]$$

1. Show that the function f^* that minimizes $\mathcal{L}(f)$ is given by $f^*(z) = \mathbb{E}[x \mid z]$.

2. Assume that the dataset consists of two points, x_1 and x_2, and the empirical distribution over the dataset is given by $p(x) = \frac{1}{2}\delta_{x_1}(x) + \frac{1}{2}\delta_{x_2}(x)$. Let the noise model be Gaussian: $q(z \mid x) = \mathcal{N}(z; x, \sigma^2 I)$. Find a closed-form expression for the optimal denoiser $f^*(z)$.