Practice Questions

azf

3 November 2023

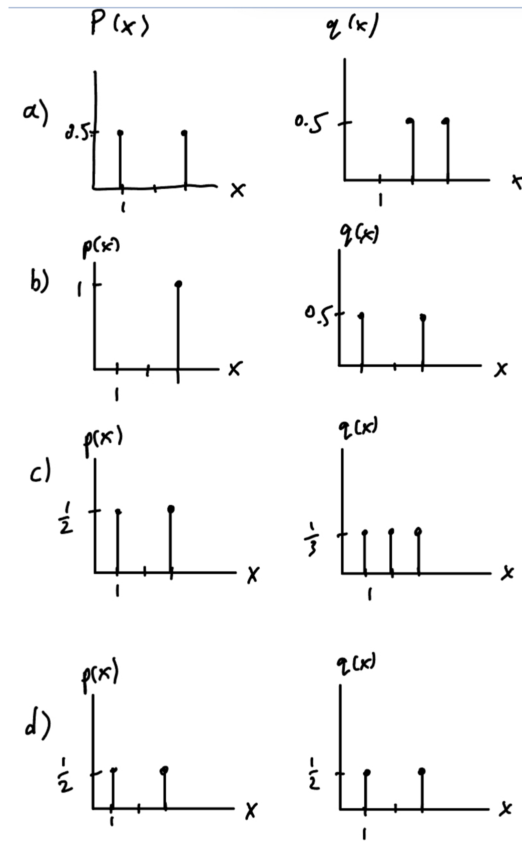
1 Kullback-Liebler Divergence

Kullback-Leibler Divergence is a measure of how far apart two distributions are:

$D(p||q) = \sum_{x \in X} p(x) \log(\frac{p(x)}{q(x)})$, where p and q are distributions defined on the same domain.

1.1

Compute $D(p||q)$ in the distributions below, using log base 2.



1.2

In general, what are the minimum and maximum of $D(p||q)$? In what situations do they occur?

2 Kalman Filtering

2.1

When using a Kalman filter, if we want to recover the maximum likelihood trajectory, can we discard the history of observations and just record the most likely states while filtering? Explain.

2.2

Consider three species U,V,W that grow independently of each other, exponentially with growth rates: U grows 2% per hour, V grows 6% per hour, and W grows 11% per hour. The goal is to estimate the initial size of each population based on the measurements of total population. Let $x_U(t)$ denote the population size of species U after t hours, for $t = 0, 1, \cdots$, and similarly for $x_V(t)$ and $x_W(t)$, so that

$x_U(t+1) = 1.02x_U(t)$, $x_V(t+1) = 1.06x_V(t)$, $x_W(t+1) = 1.11x_W(t)$.

The total population measurements are $y(t) = x_U(t) + x_V(t) + x_W(t) + v(t)$, where $v(t)$ are IID, $\mathcal{N}(0, 0.36)$.
The prior information is that $x_U(0), x_V(0), x_W(0)$ are IID $\mathcal{N}(6, 2)$ (ignore that the initial populations can be negative.).

How do you formulate this problem as a Kalman filtering problem by providing A, H, W, R?

3 Importance Sampling

Consider the probability distribution with density
$p(x, y) \propto f(x, y) = e^{-\frac{1}{2}(x^2+y^2+x^2y^2+\cos(x+.1y)+1)}$. Describe an algorithm using importance sampling to obtain samples from this distribution.

4 Particle Filtering

This is a coding problem to illustrate that particle filtering is biased for finite sample sizes.

Consider a world with four possible robot locations $X = \{x_1, x_2, x_3, x_4\}$. Initially, we draw N samples uniformly with replacement from the locations. Let Y be a binary sensor variable characterized by the following conditional probabilities:

$$p(y|x_1) = 0.8 = 1 - p(\neg y|x_1) \tag{1}$$

$$p(y|x_2) = 0.4 = 1 - p(\neg y|x_2) \tag{2}$$

$$p(y|x_3) = 0.1 = 1 - p(\neg y|x_3) \tag{3}$$

$$p(y|x_4) = 0.1 = 1 - p(\neg y|x_4) \tag{4}$$

Particle filtering uses these probabilities to generate particle weights, which are subsequently normalized and used in the resampling process. For simplicity, let us assume we only generate one new sample in the resampling process, regardless of the initial number of samples N. This sample might correspond to any of the four locations X. Thus, the sampling process defines a probability distribution over X.

4.1

What is the resulting probability distribution over X for the new sample, for $N = 1, 2, 4, 8$?

4.2

What is the KL divergence between the distributions in **4.1** and the true posterior? Observe that particle filtering is biased for finite sample sizes. Describe how particle filtering is biased, what problems it may cause, and how it changes with N.[1]

[1] 2.2, 3, 4 contain parts of problems taken from CS 287.