Practice Questions

azf

3 November 2023

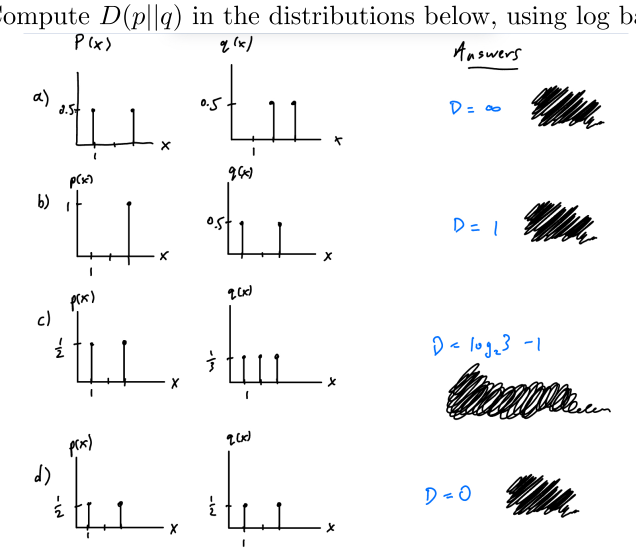
1 Kullback-Leibler Divergence

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?

Answer: $D = 0$ occurs when p=q.$D = \infty$ occurs when $\exists x. q(x) = 0, p(x) \neq 0$

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.

Answer: Yes, for Gaussians, the most probable values of the hidden variables under the marginals are equal to the most probable values under the joint. At each filtering step, we get $P(x_i|y_{0:i})$, and we are looking for $argmax_{x_{0:T}} P(x_{0:T}|y_{0:T}) \propto P(x_{0:T}, o_{0:T})$. If we don't marginalize x_{t-1} in process step and don't marginalize y_t in the measurement step, we would end up with the joint rather than the marginals.

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?

Answer: A $= [[1.02, 0, 0], [0, 1.06, 0], [0, 0, 1.11]]$, $R = 0.36$, $W = 0$, $H = [[1, 1, 1]]$

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.

Answer: Propose a q that is easy to sample (x, y) from, such as bivariate normal $e^{-\frac{1}{2}(x^2+y^2)}$. Use weight $w(x, y) = p(x, y)/q(x, y)$ for each sample.

4 Particle Filtering

This problem wasn't covered in the problem session. [1]

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