

Statistical Models with Variational Methods

Luis Antonio Ortega Andrés

November 18, 2019

1 Probability

1.1 Notation

Variables will be denoted with lower case x and a set of variables with a calligraphic symbol like \mathcal{V} . The meaning of $p(state)$ will be clear without a reference to the variable. Otherwise $p(x = state)$ will be used.

We will denote $p(x)$ the probability of x taking a specific value, this means that

$$\int_x f(x) = \int_{dom(x)} f(x = s) ds$$

1.2 Definitions

We will define some concepts from a given joint distribution $p(x, y)$, this is, the probability of two events.

Definition 1. A **marginal** $p(x)$ of the joint distribution is the distribution of a single variable given by

$$p(x) = \sum_y p(x, y) \qquad p(x) = \int_y p(x, y)$$

We can understand this as the probability of an event irrespective of the outcome of another variable.

Definition 2. The **conditional probability** of x given y is defined as

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

If $p(y) = 0$ then it is not defined. This formula is also known as **Baye's rule**. With this definition the conditional probability is the probability of one event occurring in the presence of a second event.

Now suppose we have some observed data \mathcal{D} and we want to learn about a set of parameters θ . Using Baye's rule we got that

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int_{\theta} p(\mathcal{D}|\theta)p(\theta)}$$

This shows how from a *generative model* $p(\mathcal{D}|\theta)$ of the dataset and a *prior belief* $p(\theta)$, we can infer the *posterior* distribution $p(\theta|\mathcal{D})$.

Example. Consider a study where the relation of a disease D and an habit H is beeing investigated. Consider $p(D) = 10^{-5}$, $p(H) = 0.5$ and $p(H|D) = 0.9$. What is the probability that a person with habit H will have desease D ?

$$p(D|S) = \frac{p(D, H)}{p(D)} = \frac{p(H|D)p(D)}{p(H)} = \frac{0.9 \times 10^{-5}}{0.5} = 1.8 \times 10^{-5}$$

If we set the probability of having habit H to a much lower value as $p(H) = 0.001$, then the above calcu-
lation gives aproximately 1/100.

Intuitively, a smaller number of people have the habit and most of them have the desease. This means
that the relation between having the desease and the habit is stronger.

2 Graphical Models

