Matthew N. Bernstein
Computational Biologist at Immunitas Therapeutics

Follow

The evidence lower bound (ELBO)

3 minute read

Published: May 25, 2020

The evidence lower bound is an important quantity at the core of a number of important algorithms used in statistical inference including expectation-maximization and variational inference. In this post, I describe its context, definition, and derivation.

Introduction

The **evidence lower bound (ELBO)** is an important quantity that lies at the core of a number of important algorithms in probabilistic inference such as expectation-maximization and variational infererence. To understand these algorithms, it is helpful to understand the ELBO.

Before digging in, let's review the probabilistic inference task for a latent variable model. In a latent variable model, we posit that our observed data x is a realization from some random variable X. Moreover, we posit the existence of another random variable Z where X and Z are distributed according to a joint distribution $p(X, Z; \theta)$ where θ parameterizes the distribution. Unfortunately, our data is *only* a realization of X, not Z, and therefore Z remains unobserved (i.e. latent).

There are two predominant tasks that we may be interested in accomplishing:

1. Given some fixed value for θ, compute the posterior distribution $p(Z \mid X; \theta)$
2. Given that θ is unknown, find the maximum likelihood estimate of θ:

$$\text{argmax}_\theta l(\theta)$$

where $l(\theta)$ is the log-likelihood function:

$$l(\theta) := \log p(x; \theta) = \log \int_z p(x, z; \theta)\, dz$$

Variational inference is used for Task 1 and expectation-maximization is used for Task 2. Both of these algorithms rely on the ELBO.

What is the ELBO?

To understand the evidence lower bound, we must first understand what we mean by "evidence". The **evidence**, quite simply, is just a name given to the likelihood function evaluated at a fixed θ:

$$\text{evidence} := \log p(x; \theta)$$

Why is this quantity called the "evidence"? Intuitively, if we have chosen the right model p and θ, then we would expect that the marginal probability of our observed data x, would be high. Thus, a higher value of $\log p(x; \theta)$ indicates, in some sense, that we may be on the right track with the model p and parameters θ that we have chosen. That is, this quantity is "evidence" that we have chosen the right model for the data.

If we happen to also know (or posit) that Z follows some distribution denoted by q (and that $p(x, z; \theta) := p(x \mid z; \theta)q(z)$), then the evidence lower bound is, well, just a lower bound on the evidence that makes use of the known (or posited) q. Specifically,

$$\log p(x; \theta) \geq E_{Z \sim q}\left[\log \frac{p(x, Z; \theta)}{q(Z)}\right]$$

where the ELBO is simply the right-hand side of the above inequality:

$$ELBO := E_{Z \sim q}\left[\log \frac{p(x, Z; \theta)}{q(Z)}\right]$$

Derivation

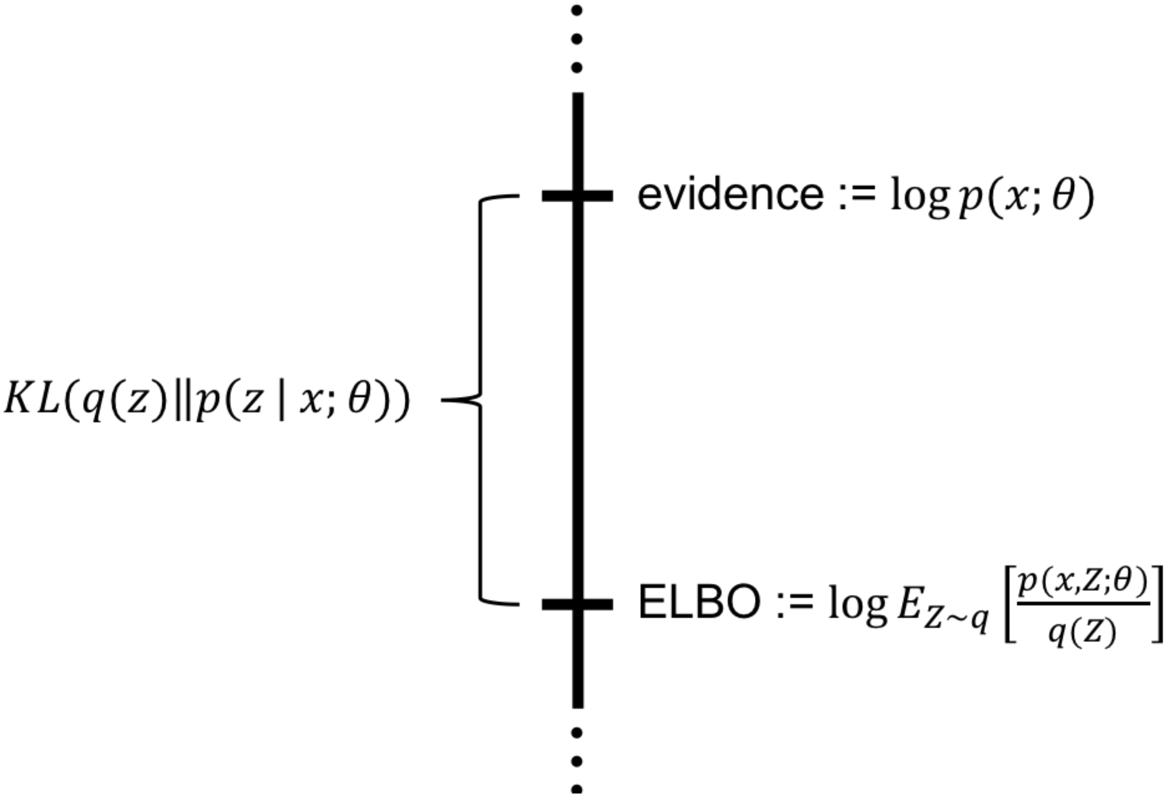
We derive this lower bound as follows:

$$\begin{aligned}
\log p(x; \theta) &= \log \int p(x, z; \theta)\, dz \\
&= \log \int p(x, z; \theta) \frac{q(z)}{q(z)}\, dz \\
&= \log E_{Z \sim q}\left[\frac{p(x, Z)}{q(z)}\right] \\
&\geq E_{Z \sim q}\left[\log \frac{p(x, Z)}{q(z)}\right]
\end{aligned}$$

This final inequality follows from Jensen's Inequality.

The gap between the evidence and the ELBO

It turns out that the gap between the evidence and the ELBO is precisely the Kullback-Leibler divergence between $p(z \mid x; \theta)$ and $q(z)$. This fact forms the basis of the variational inference algorithm for approximate Bayesian inference!



This can be derived as follows:

$$
\begin{aligned}
KL(q(z) \ \ || \ p(z \mid x; \theta)) &:= E_{Z \sim q}\left[\log \frac{q(Z)}{p(Z \mid x; \theta)}\right] \\
&= E_{Z \sim q}\left[\log q(Z)\right] - E_{Z \sim q}\left[\log \frac{p(x, Z; \theta)}{p(x; \theta)}\right] \\
&= E_{Z \sim q}\left[\log q(Z)\right] - E_{Z \sim q}\left[\log p(x, Z; \theta)\right] + E_{Z \sim q}\left[\log p(x; \theta)\right] \\
&= \log p(x; \theta) - E_{Z \sim q}\left[\log \frac{p(x, Z; \theta)}{q(z)}\right] \\
&= \text{evidence} - \text{ELBO}
\end{aligned}
$$

Tags: evidence lower bound | machine learning | probability | statistics | tutorial

Previous | Next