Testing ADVI

Alp Kucukelbir

April 7, 2015

I document how to calculate ADVI quantities (objective function, gradients) in closed form for a few simple models. These are useful for testing.

1 Univariate Model without Constraints

Consider the following simple model.

$$p(\theta) = \mathcal{N}(\theta \,;\, \mu_0, \sigma_0^2)$$
$$p(y_i \mid \theta) = \mathcal{N}(y_i \,;\, \theta, \sigma^2)$$

The joint distribution is

$$p(y, \theta) = \mathcal{N}(\theta \,;\, \mu_0, \sigma_0^2) \prod_{i=1}^{n} \mathcal{N}(\theta \,;\, y_i, \sigma^2)$$

where we write the likelihoods as functions of the latent variable θ.

Now consider an approximating Gaussian distribution

$$q(\theta) = \mathcal{N}(\theta \,;\, \mu_q, \sigma_q^2).$$

The *evidence lower bound* (ELBO) is

$$\mathcal{L} = \mathbb{E}_q[\log p(y, \theta)] + \mathbb{E}_q[-\log q(\theta)].$$

The joint distribution. We can write the joint distribution as a scaled Gaussian

$$p(y, \theta) = S_J \mathcal{N}(\theta \,;\, \mu_J, \sigma_J^2)$$

where

$$\frac{1}{\sigma_J^2} = \frac{1}{\sigma_0^2} + \sum_{i=1}^{n} \frac{1}{\sigma^2}$$
$$\mu_J = \left(\frac{\mu_0}{\sigma_0^2} + \sum_{i=1}^{n} \frac{y_i}{\sigma^2} \right) \sigma_J^2$$
$$S_J = \frac{1}{(2\pi)^{n/2}} \sqrt{\frac{\sigma_J^2}{\sigma_0^2 \prod_{i=1}^{n} \sigma^2}} \exp\left[-\frac{1}{2} \left(\frac{\mu_0^2}{\sigma_0^2} + \sum_{i=1}^{n} \frac{y_i^2}{\sigma^2} - \frac{\mu_J^2}{\sigma_j^2} \right) \right]$$

These equations are slightly modified from (Bromiley, 2003).

The ELBO. We can analytically write the ELBO as

$$\begin{aligned}
\mathscr{L} &= \log S_J + \log \frac{1}{\sqrt{2\pi\sigma_J^2}} - \frac{1}{2}\mathbb{E}_q\left[\frac{(\theta-\mu_J)^2}{\sigma_J^2}\right] + \frac{1}{2}\left(1+\log 2\pi + \log \sigma_q^2\right) \\
&= \log S_J + \log \frac{1}{\sqrt{2\pi\sigma_J^2}} - \frac{1}{2}\left(\frac{(\mu_J-\mu_q)^2}{\sigma_J^2} + \frac{\sigma_q^2}{\sigma_J^2}\right) + \frac{1}{2}\left(1+\log 2\pi + \log \sigma_q^2\right)
\end{aligned}$$

The quadratic expectation is easy (Roweis, 1999).

2 Univariate Model with Constraints

Consider the above model such that we perform computations in the transformed space

$$\theta^\dagger = \log\theta.$$

The joint distribution becomes

$$\begin{aligned}
p(y,\theta) &= p(y, \exp\theta^\dagger)\cdot\left|\det J_{\exp}(\theta^\dagger)\right| \\
&= p(y,\exp\theta^\dagger)\exp\theta^\dagger.
\end{aligned}$$

The approximation is in the transformed space

$$q(\theta^\dagger) = \mathcal{N}(\theta\,;\,\mu_q,\sigma_q^2).$$

The ELBO thus becomes

$$\mathscr{L} = \mathbb{E}_q[\log p(y,\exp\theta^\dagger) + \theta^\dagger] + \mathbb{E}_q[-\log q(\theta^\dagger)].$$

This is a bit trickier. We need the moment generating function of a Gaussian to evaluate everything.

$$\begin{aligned}
\mathscr{L} &= \log S_J + \log\frac{1}{\sqrt{2\pi\sigma_J^2}} - \frac{1}{2\sigma_J^2}\mathbb{E}_q\left[(\exp\theta^\dagger - \mu_J)^2\right] + \mathbb{E}_q[\theta^\dagger] + \frac{1}{2}\left(1+\log 2\pi+\log\sigma_q^2\right) \\
&= \log S_J + \log\frac{1}{\sqrt{2\pi\sigma_J^2}} - \frac{1}{2\sigma_J^2}\left[\exp 2\mu_q \exp 2\sigma_q^2 - 2\mu_J \exp\mu_q \exp\frac{1}{2}\sigma_q^2 + \mu_J^2\right] \\
&\quad + \mu_q + \frac{1}{2}\left(1+\log 2\pi + \log\sigma_q^2\right)
\end{aligned}$$

3 Multivariate Model without Constraints

Consider the following d-dimensional model.

$$\begin{aligned}
p(\theta) &= \mathcal{N}(\theta\,;\,\mu_0, I_{d\times d}) \\
p(y_i \mid \theta) &= \mathcal{N}(y_i\,;\,\theta, I_{d\times d})
\end{aligned}$$

The joint distribution is

$$p(y,\theta) = \mathcal{N}(\theta\,;\,\mu_0, I_{d\times d})\prod_{i=1}^{n}\mathcal{N}(\theta\,;\,y_i, I_{d\times d})$$

where we write the likelihoods as functions of the latent variable θ.

Now consider an approximating Gaussian distribution

$$q(\theta) = \mathcal{N}(\theta\,;\,\mu_q, I_{d\times d}).$$

The ELBO is

$$\mathcal{L} = \mathbb{E}_q[\log p(y,\theta)] + \mathbb{E}_q[-\log q(\theta)].$$

The log-joint distribution. We can write the log-joint distribution as

$$\log p(y,\theta) = \zeta + \eta_J^\top \theta - \frac{1}{2}\theta^\top \Lambda_J \theta$$

where

$$\begin{aligned}
\zeta &= -\frac{1}{2}\left((n+1)d\log 2\pi + \mu_0^\top \mu_0 + \sum_i y_i^\top y_i\right)\\
\eta_J &= \mu_0 + \sum_i y_i\\
\Lambda_J &= (n+1) I_{d\times d}
\end{aligned}$$

These equations are slightly modified from (Bromiley, 2003).

The ELBO. We can analytically write the ELBO as

$$\begin{aligned}
\mathcal{L} &= \zeta + \eta_J^\top \mathbb{E}_q[\theta] - \frac{1}{2}\mathbb{E}_q\left[\theta^\top \Lambda_J \theta\right] + \frac{1}{2}d\log 2\pi\\
&= \zeta + \eta_J^\top \mu_q - \frac{1}{2}\left(3\mu_q^\top \mu_q^\top + \mathrm{Tr}\Lambda_J\right) + \frac{1}{2}d\log 2\pi\\
&= \zeta + \eta_J^\top \mu_q - \frac{1}{2}\left(3\mu_q^\top \mu_q^\top + 3d\right) + \frac{1}{2}d(1+\log 2\pi)
\end{aligned}$$

The expectations are from (Roweis, 1999).

4 Multivariate Model with Constraints

Consider the above model such that we perform computations in the transformed space

$$\theta^\dagger = \log\theta.$$

The logjoint distribution becomes

$$\begin{aligned}
\log p(y,\theta) &= p(y, \exp\theta^\dagger) + \log\left|\det J_{\exp}(\theta^\dagger)\right|\\
&= p(y, \exp\theta^\dagger) + \sum_d \theta_d^\dagger.
\end{aligned}$$

The approximation is in the transformed space

$$q(\theta^\dagger) = \mathcal{N}(\theta\,;\,\mu_q, I_{d\times d}).$$

The ELBO thus becomes

$$\mathcal{L} = \mathbb{E}_q[\log p(y, \exp\theta^\dagger) + \sum_d \theta_d^\dagger] + \mathbb{E}_q[-\log q(\theta^\dagger)].$$

This is a bit trickier. We need the moment generating function of a Gaussian to evaluate everything.

$$\mathcal{L} = \zeta + \eta_J^\top \exp(\mu_q + 0.5) - \frac{1}{2}(n+1)\sum_d \exp(2\mu_d + 1) + \sum_d \mu_d + \frac{1}{2}d(1+\log 2\pi)$$

References

Bromiley, P. (2003). Products and convolutions of Gaussian probability density functions. Technical report, Tina-Vision Memo. http://tina.wiau.man.ac.uk/docs/memos/2003-003.pdf.

Roweis, S. (1999). Gaussian identities. Technical report, University of Toronto.