Local Global Approximations (working title)

John T. Ormerod

October 4, 2022

1 Introduction

We will continue to work on the Lasso becuse that is a good place to start.

2 The Lasso Distribution

If $x \sim \text{Lasso}(a, b, c)$ with then it has density given by

$$p(x, a, b, c) = Z^{-1} \exp\left(-\tfrac{1}{2}ax^2 + bx - c|x|\right)$$

where $x \in \mathbb{R}$, $a > 0$, $b \in \mathbb{R}$, $c > 0$, and Z is the normalizing constant. Then

$$\begin{aligned}
Z(a,b,c) &= \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}ax^2 + bx - c|x|\right] dx \\
&= \int_{0}^{\infty} \exp\left[-\tfrac{1}{2}ax^2 + (b-c)x\right] dx + \int_{-\infty}^{0} \exp\left[-\tfrac{1}{2}ax^2 + (b+c)x\right] dx \\
&= \int_{0}^{\infty} \exp\left[-\tfrac{1}{2}ax^2 + (b-c)x\right] dx + \int_{0}^{\infty} \exp\left[-\tfrac{1}{2}ay^2 - (b+c)y\right] dy \\
&= \int_{0}^{\infty} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2} + \frac{\mu_1^2}{2\sigma^2}\right] dx + \int_{0}^{\infty} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma^2} + \frac{\mu_2^2}{2\sigma^2}\right] dy \\
&= \sqrt{2\pi\sigma^2}\left[\exp\left\{\frac{\mu_1^2}{2\sigma^2}\right\}\int_0^\infty \phi(x;\mu_1,\sigma^2)dx + \exp\left\{\frac{\mu_2^2}{2\sigma^2}\right\}\int_0^\infty \phi(y;\mu_2,\sigma^2)dy\right] \\
&= \sqrt{2\pi\sigma^2}\left[\exp\left\{\frac{\mu_1^2}{2\sigma^2}\right\}\{1-\Phi(-\mu_1/\sigma)\} + \exp\left\{\frac{\mu_2^2}{2\sigma^2}\right\}\{1-\Phi(-\mu_2/\sigma)\}\right] \\
&= \sqrt{2\pi\sigma^2}\left[\exp\left(\frac{\mu_1^2}{2\sigma^2}\right)\Phi\left(\frac{\mu_1}{\sigma}\right) + \exp\left(\frac{\mu_2^2}{2\sigma^2}\right)\Phi\left(\frac{\mu_2}{\sigma}\right)\right] \\
&= \sigma\left[\frac{\Phi(\mu_1/\sigma)}{\phi(\mu_1/\sigma)} + \frac{\Phi(\mu_2/\sigma)}{\phi(\mu_2/\sigma)}\right]
\end{aligned}$$

where $\mu_1 = (b-c)/a$, $\mu_2 = -(c+b)/a$ and $\sigma^2 = 1/a$. Care should be taken when evaluating Z (which is prone to overflow and divide by zero problems) and is a function of the Mills ratio.

2.1 The Moment Generating function and Moments

The moment generating function requires almost identical calculations with b replaced with $b+t$.

$$M(t) = \frac{Z(a,b+t,c)}{Z(a,b,c)}$$

While this is true it doesn't look useful for calculating moments.

2.2 Moments

The moments of the Lasso distribution are:

$$\begin{aligned}
E(x^r) &= Z^{-1}\int_{-\infty}^{\infty} x^r \exp\left[-\tfrac{1}{2}ax^2 + bx - c|x|\right] dx \\
&= Z^{-1}\int_{0}^{\infty} x^r \exp\left[-\tfrac{1}{2}ax^2 + (b-c)x\right] dx + \int_{-\infty}^{0} x^r \exp\left[-\tfrac{1}{2}ax^2 + (b+c)x\right] dx \\
&= Z^{-1}\int_{0}^{\infty} x^r \exp\left[-\tfrac{1}{2}ax^2 + (b-c)x\right] dx + (-1)^r\int_{0}^{\infty} y^r \exp\left[-\tfrac{1}{2}ay^2 - (b+c)y\right] dy \\
&= Z^{-1}\sqrt{2\pi\sigma^2}\exp\left(\frac{\mu_1^2}{2\sigma^2}\right)\int_0^\infty x^r\phi(x;\mu_1,\sigma^2)dx \\
&\qquad +(-1)^r\sqrt{2\pi\sigma^2}\exp\left(\frac{\mu_2^2}{2\sigma^2}\right)\int_0^\infty y^r\phi(y;\mu_2,\sigma^2)dy \\
&= \frac{\sigma}{Z}\left[\frac{\Phi(\mu_1/\sigma)}{\phi(\mu_1/\sigma)}\frac{\int_0^\infty x^r\phi(x;\mu_1,\sigma^2)dx}{\Phi(\mu_1/\sigma)} + (-1)^r\frac{\Phi(\mu_2/\sigma)}{\phi(\mu_2/\sigma)}\frac{\int_0^\infty y^r\phi(y;\mu_2,\sigma^2)dy}{\Phi(\mu_2/\sigma)}\right] \\
&= \frac{\sigma}{Z}\left[\frac{\Phi(\mu_1/\sigma)}{\phi(\mu_1/\sigma)}\mathbb{E}(A^r) + (-1)^r\frac{\Phi(\mu_2/\sigma)}{\phi(\mu_2/\sigma)}\mathbb{E}(B^r)\right]
\end{aligned}$$

where $A \sim TN_+(\mu_1,\sigma^2)$, $B \sim TN_+(\mu_2,\sigma^2)$ and TN_+ is denotes the positively truncated normal distribution. Note that

$$\mathbb{E}(A) = \mu_1 + \frac{\sigma\phi(\mu_1/\sigma)}{\Phi(\mu_1/\sigma)} = \mu_1 + \sigma\zeta_1(\mu_1/\sigma)$$

and

$$\mathbb{V}(A) = \sigma^2\left[1+\zeta_2(\mu_1/\sigma)\right]$$

where $\zeta_k(x) = d^k\log\Phi(x)/dx^k$, $\zeta_1(t) = \phi(t)/\Phi(t)$, $\zeta_2(t) = -t\,\zeta_1(t) - \zeta_1(t)^2$. Here $\zeta_1(x)$ is the inverse Mills ratio which too needs to be treated with care. Hence,

$$\mathbb{E}(A^2) = \mathbb{V}(A) + \mathbb{E}(A)^2 = \sigma^2\left[1+\zeta_2(\mu_1/\sigma)\right] + \left[\mu_1 + \sigma\zeta_1(\mu_1/\sigma)\right]^2$$

We now have sufficient information to calculate the moments of the Lasso distribution. We also have sufficient information to implement a VB approximation.

2.3 CDF

Similarly if $z \leq 0$ the CDF is given by

$$\begin{aligned}
P(Z < z) &= Z^{-1} \int_{-\infty}^{z} \exp\left[-\tfrac{1}{2}ax^2 + (b+c)x\right] dx \\
&= Z^{-1}\sqrt{2\pi\sigma^2} \exp\left(\frac{\mu_2^2}{2\sigma^2}\right) \int_{-\infty}^{z} \phi(x; -\mu_2, \sigma^2) \\
&= Z^{-1}\sqrt{2\pi\sigma^2} \exp\left(\frac{\mu_2^2}{2\sigma^2}\right) \Phi\left(\frac{z+\mu_2}{\sigma}\right) \\
&= \frac{\sigma}{Z} \frac{\Phi\left(\frac{z+\mu_2}{\sigma}\right)}{\phi(\mu_2/\sigma)}
\end{aligned}$$

and if $z > 0$ we have

$$\begin{aligned}
P(Z < z) &= Z^{-1} \int_{-\infty}^{z} \exp\left(-\tfrac{1}{2}ax^2 + bx - c|x|\right) dx \\
&= Z^{-1}\sqrt{2\pi\sigma^2} \left[\exp\left(\frac{\mu_1^2}{2\sigma^2}\right) \int_{0}^{z} \phi(x; \mu_1, \sigma^2) dx + \exp\left(\frac{\mu_2^2}{2\sigma^2}\right) \Phi\left(\frac{\mu_2}{\sigma}\right)\right] \\
&= \frac{\sigma}{Z}\left[\frac{\Phi\left(\frac{z-\mu_1}{\sigma}\right) - \Phi\left(\frac{-\mu_1}{\sigma}\right)}{\phi(\mu_1/\sigma)} + \frac{\Phi\left(\mu_2/\sigma\right)}{\phi(\mu_2/\sigma)}\right]
\end{aligned}$$

2.4 Inverse CDF

For the inverse CDF we again have two cases. Let $u = P(Z < z)$. When

$$u \leq \frac{\sigma}{Z} \frac{\Phi\left(\mu_2/\sigma\right)}{\phi(\mu_2/\sigma)}$$

we solve

$$u = \frac{\sigma\Phi\left(\frac{z+\mu_2}{\sigma}\right)}{Z\phi(\mu_2/\sigma)}$$

for z to obtain

$$z = \mu_2 + \sigma\,\Phi^{-1}\left[(Z/\sigma)\phi_\sigma(\mu_2/\sigma)u\right]$$

When

$$u > \frac{\sigma}{Z} \frac{\Phi\left(\mu_2/\sigma\right)}{\phi(\mu_2/\sigma)}$$

we need to solve

$$u = \frac{\sigma}{Z}\left[\frac{\Phi\left(\frac{z-\mu_1}{\sigma}\right) - \Phi\left(\frac{-\mu_1}{\sigma}\right)}{\phi(\mu_1/\sigma)} + \frac{\Phi\left(\mu_2/\sigma\right)}{\phi(\mu_2/\sigma)}\right]$$

for z to obtain

$$z = \mu_1 + \sigma\,\Phi^{-1}\left[\phi(\mu_1/\sigma)\left\{\frac{Zu}{\sigma} - \frac{\Phi\left(\mu_2/\sigma\right)}{\phi(\mu_2/\sigma)}\right\} + \Phi\left(\frac{-\mu_1}{\sigma}\right)\right]$$

which also involves the Mills ratio.

3 Tasks

Suppose that conformably we have

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $q(\boldsymbol{\theta}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ approximates $p(\boldsymbol{\theta} \mid \mathcal{D})$.

1. Verify the algebra in Section 2 above for the Lasso distribution.

2. Show that
$$p(\boldsymbol{\theta}_1 \mid \mathcal{D}) = \int p(\boldsymbol{\theta}_1 \mid \mathcal{D}, \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2 \mid \mathcal{D}) d\boldsymbol{\theta}_2$$

3. Find $q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1)$.

4. Suppose that we approximate $p(\boldsymbol{\theta}_2 \mid \mathcal{D})$ by $q(\boldsymbol{\theta}_2) = N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. Suppose we use this to approximate $p(\boldsymbol{\theta}_1 \mid \mathcal{D})$ by
$$q^*(\boldsymbol{\theta}_1) = \int p(\boldsymbol{\theta}_1 \mid \mathcal{D}, \boldsymbol{\theta}_2) q(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2$$
Suppose that the mean of $q^*(\boldsymbol{\theta}_1)$ is $\boldsymbol{\mu}_1^*$, covariance is $\boldsymbol{\Sigma}_{11}^*$, and $q^*(\boldsymbol{\theta}_1) \approx N(\boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_{11}^*)$. We update $q(\boldsymbol{\theta})$ via
$$q^*(\boldsymbol{\theta}) = q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1) \phi(\boldsymbol{\theta}_1; \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_{11}^*)$$
Find $q^*(\boldsymbol{\theta})$.

5. Consider the multivariate Lasso distribution
$$p(\mathbf{x}) \propto \exp\left(-\tfrac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} - c\|\mathbf{x}\|_1\right)$$
where $\mathbf{A} \in \mathcal{S}_d^+$ is a positive definite matrix of dimension d, $\mathbf{b} \in \mathbb{R}^2$ and $c > 0$ with $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^d |x_j|$. For the case $d = 2$
    - Find the normalizing constant.
    - Find an expression for the expectation and the covariance.