Count Data Issues with Integrated Nested Laplace Approximations

Adam Walder

April 23^{rd}, 2018

Latent Gaussian Models consist of three layers.

- Likelihood: $\boldsymbol{y}|\boldsymbol{\eta},\boldsymbol{\theta} \sim \prod_i p(y_i|\eta_i,\boldsymbol{\theta})$
- Latent Field: $\boldsymbol{\eta}|\boldsymbol{\theta} \sim p(\boldsymbol{\eta}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0},\Sigma(\theta))$
- Hyper priors: $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$

A latent Gaussian field with a sparse precision matrix, Q,

$$\boldsymbol{\eta}|\boldsymbol{\theta} \sim p(\boldsymbol{\eta}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0},Q(\boldsymbol{\theta})^{-1})$$

is referred to as a **Gauss-Markov Random Field.**

- Benefit of sparse Q.
- Why LGMs?

- **Why do we need INLA?**
  - MCMC alternative
- **What are we after with INLA?**
  - $\pi(\eta_i|\boldsymbol{y})$ and $\pi(\theta_i|\boldsymbol{y})$
- **When is INLA applicable?**
  - $|\boldsymbol{\theta}|$ is small.
  - $\boldsymbol{\eta}|\boldsymbol{\theta}$ is GMRF when dimension is high.
  - Each y_i depends on only one η_i.
- **Why is it called INLA?**
  1. Integrated: Numerical integration gives our estimates.
  2. Nested: We use an approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ to get $\pi(\eta_i|\boldsymbol{y})$
  3. LA: Used in approximating $\pi(\boldsymbol{\theta}|\boldsymbol{y})$

Goal: Obtain marginal distributions $\pi(\theta_i|\boldsymbol{y})$ and $\pi(\eta_i|\boldsymbol{y})$

Method: We obtain the marginals through numerical integration

$$\begin{aligned}
\pi(\theta_i|\boldsymbol{y}) &\approx \sum_k \tilde{\pi}(\boldsymbol{\theta}_k^{(i)}|\boldsymbol{y}) \text{ x } \Delta(\boldsymbol{\theta}_k^{(-i)}) \\
\pi(\eta_i|\boldsymbol{y}) &\approx \sum_k \tilde{\pi}(\boldsymbol{\eta}_i|\boldsymbol{\theta}_k, \boldsymbol{y}) \text{ x } \tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y}) \text{ x } \Delta(\boldsymbol{\theta}_k)
\end{aligned}$$

Preliminary Tasks:

1. Obtain approximation $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$
2. Explore $\log(\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}))$ to obtain $\Delta(\boldsymbol{\theta})$
3. Obtain approximation $\tilde{\pi}(\eta_i|\boldsymbol{\theta}, \boldsymbol{y})$

STEP 1: FIND $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$

The approximation for the marginal of the parameters is given by

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*(\boldsymbol{\theta})} \quad (1)$$

The Gaussian approximation in the denominator must be found as follows,

$$\tilde{\pi}(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{\theta}) \propto exp(-\frac{1}{2}\boldsymbol{\eta}' Q(\boldsymbol{\theta})\boldsymbol{\eta} + \sum \log(\pi(y_i|\boldsymbol{\theta}, \eta_i))) \quad (2)$$

We expand a second-order Taylor expansion about an initial modal guess $\boldsymbol{\mu}^{(0)}$. The result is

$$\log(\pi(y_i|\boldsymbol{\theta}, \eta_i)) \approx b_i(\mu_i^{(k)})\eta_i - \frac{1}{2}c_i(\mu_i^{(k)})\eta_i^2$$

To finish finding $\boldsymbol{\eta}^*(\boldsymbol{\theta})$ the following NR algorithm is used,

$$\boldsymbol{\eta}^{(k+1)} = (Q(\boldsymbol{\theta}) + diag(\boldsymbol{c}(\boldsymbol{\eta}^{(k)})))^{-1}\boldsymbol{b}(\boldsymbol{\eta}^{(k)})$$

The result produces the following Gaussian approximation to the posterior distribution

$$\pi_G(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{y}) \sim \mathcal{N}(\boldsymbol{\eta}^*(\boldsymbol{\theta}), [Q(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{c}(\boldsymbol{\eta}^*(\boldsymbol{\theta})))]^{-1})$$

Note: Still have a *GMRF*.
Note: The algorithm involves iteratively solving a system.

Step 2: Explore $\log(\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}))$

(1) Find mode:

$$\boldsymbol{\theta}^* = \text{argmax}_{\boldsymbol{\theta}} \log(\tilde{\pi}(\theta|\boldsymbol{y}))$$

(2) z-parameterization

$$\Sigma = H^{-1} = V\Lambda^{1/2}V'$$

Define: $\theta(\boldsymbol{z}) = \boldsymbol{\theta}^* + V\Lambda^{1/2}\boldsymbol{z}$

(3) Define δ_z and δ_π. Let k = dim($\boldsymbol{\theta}$)

for(i in 1:k)

$\boldsymbol{z} = (0, , .., 0)$ $\quad$ $z_i = \delta_z$

while($|\log(\tilde{\pi}(\boldsymbol{\theta}(\mathbf{0})|\boldsymbol{y}) - \log(\tilde{\pi}(\boldsymbol{\theta}(\boldsymbol{z})|\boldsymbol{y})| < \delta_\pi$)

store $\boldsymbol{\theta}(\boldsymbol{z})$ and $\log(\tilde{\pi}(\boldsymbol{\theta}(\boldsymbol{z})|\boldsymbol{y})$

$z_i = z_i + \delta_z$

Repeat for the opposite direction.

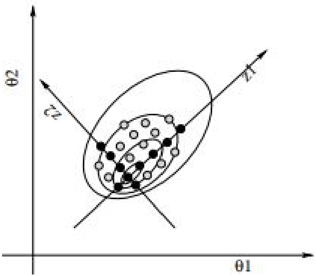
We also need to compute all the intermediate combinations. In dimensions > 2, Box and Behnken design is used.

We can approximate the density as follows

$$\tilde{\pi}(\eta_i|\boldsymbol{\theta}, \boldsymbol{y}) \sim \mathcal{N}(\mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta}))$$

- We already have $\mu_i(\boldsymbol{\theta})$ from step 2.
- We only need to obtain the marginal variances.
- Other options are LA and SLA.

Complete the Integration



Generalized Latent Gaussian Model

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) &\sim x_i\beta + \eta_i \\
L\boldsymbol{\eta} &\sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 \boldsymbol{I})
\end{aligned}
$$

The priors, $\boldsymbol{\theta} = (\beta, \sigma_\eta^2)$, are given by

$$
\begin{aligned}
\beta &\sim \mathcal{N}(0, \sigma_\beta^2) \\
\sigma_\eta^2 &\sim I.G.(u, v)
\end{aligned}
$$

- L = Q + κ^2I.
- Q is a finite difference matrix on a 2-D unit square.

Note: L*L is sparse, so we are dealing with a **GMRF**.

- The algorithm fails in the first step.
- We can not obtain a Gaussian approximation to the denominator.

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*(\boldsymbol{\theta})}$$

The NR algorithm is very sensitive to the initial points of optim.

Recall: NR requires iterating until convergence

$$\boldsymbol{\mu}^{(k+1)} = [\frac{1}{\sigma_\eta^2} L^2 + \mathrm{diag}(\boldsymbol{c}^{(k)})]^{-1} \boldsymbol{b}^{(k)}$$

$$b_i^{(k)} = y_i - exp(x_i \beta + \mu_i^{(k)}) \quad \text{and} \quad c_i^{(k)} = -exp(x_i \beta + \mu_i^{(k)})$$

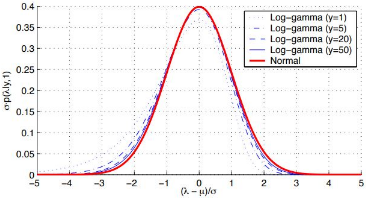
Normal Approximation to the Log-Gamma Distribution

The Log-Gamma distribution

$$\pi(v|a,b) = \frac{1}{\Gamma(a)b^a} e^{va} e^{-\frac{e^v}{b}} \approx \mathcal{N}(\log(ab), a^{-1})$$

We can now write rewrite $\pi(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})$ as follows,

$$[\boldsymbol{y}|\boldsymbol{\lambda}] = \prod \frac{1}{y_i!} \lambda_i^{y_i} e^{-\lambda_i} = \prod \frac{1}{y_i} [\frac{1}{(y_i-1)!} e^{\log(\lambda_i) y_i} e^{-e^{\log(\lambda_i)}}] \approx \prod \frac{1}{y_i} \mathcal{N}(\log(y_i), y_i^{-1})$$



The Gaussian approximation is now,

$$\begin{aligned}\tilde{\pi}(\boldsymbol{\eta}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \exp(-\frac{1}{2}\boldsymbol{\eta}' Q(\boldsymbol{\theta})\boldsymbol{\eta} + \sum_i \log(\pi(y_i|\eta_i,\boldsymbol{\theta}))) \\ &\propto \exp(-\frac{1}{2}\boldsymbol{\eta}' Q(\boldsymbol{\theta})\boldsymbol{\eta} + \sum_i \log(\mathcal{N}_{\log(\lambda_i)}(\log(y_i), y_i^{-1})))\end{aligned}$$

Now we have

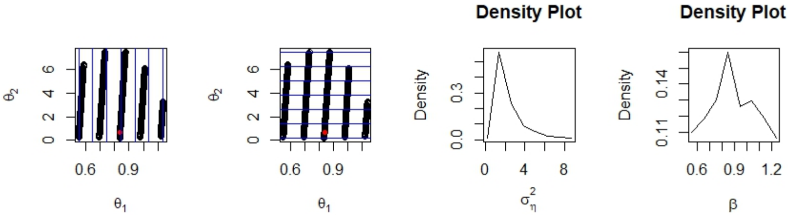
$$\tilde{\pi}(\boldsymbol{\eta}|\boldsymbol{y},\boldsymbol{\theta}) \sim \mathcal{N}(Q^*(\boldsymbol{\theta})^{-1}(\log(\boldsymbol{y}) - \boldsymbol{x}'\beta), Q^*(\boldsymbol{\theta})^{-1})$$

Where,

$$Q^*(\boldsymbol{\theta}) = \frac{1}{\sigma_\eta^2} L^2 + \text{diag}(\boldsymbol{y})$$

What was accomplished?

Results



$\hat{\beta}$	95% CI: β	$\hat{\sigma}_\eta^2$	95% CI: σ_η^2
0.8479	(0.8896, 0.9314)	2.3889	(1.9000, 2.8778)

Concluding Thoughts

- **Topics Unmentioned**
  - INLA package, Issues extending to $|\boldsymbol{\theta}| > 2$
- **Criticisms**
  - Lack of Clarity by authors
- **Caveat of "blackboxing"**
  - No justification/intuition for subtle decisions
- **Future Work**
  - Interested in how they chose δ_π, find new interpolant.