MAST90125: Bayesian Statistical Learning

Lecture 15: Extending linear regression

Feng Liu and Guoqi Qian

What have we learned so far

- In the previous lecture, we learned Bayesian computing and inference for $\boldsymbol{\beta}$ and σ^2 in linear regression analysis.
- In this lecture, we will learn Bayesian computing and inference for $\boldsymbol{\beta}$ in the following models
  - Ridge regression/random effects regression/mixed model
  - LASSO

Normal priors for $\boldsymbol{\beta}$

- Let's start from a general case in which $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2_{\boldsymbol{\beta}}\mathbf{K})$, where $\mathbf{K}$ is an arbitrary variance-covariance matrix. We will then deduce results for various special cases.
- To make derivations easier, we will work with the precision or inverse variances ($\tau = (\sigma^2)^{-1}$) rather than variances, and assume the inverse variances *a priori* are distributed $\text{Ga}(\alpha_i, \gamma_i)$. Further, we assume that $\mathbf{K}$ is known and does not need to be estimated.

Normal priors for $\boldsymbol{\beta}$

- Let's start from a general case in which $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2_{\boldsymbol{\beta}}\mathbf{K})$, where $\mathbf{K}$ is an arbitrary variance-covariance matrix. We will then deduce results for various special cases.
- To make derivations easier, we will work with the precision or inverse variances ($\tau = (\sigma^2)^{-1}$) rather than variances, and assume the inverse variances *a priori* are distributed $\text{Ga}(\alpha_i, \gamma_i)$. Further, we assume that $\mathbf{K}$ is known and does not need to be estimated.
- How many priors do we consider here?

Priors and likelihood

- Given a normal regression model, the likelihood $p(\mathbf{y}|\boldsymbol{\beta}, \tau_e, \mathbf{X})$ is

$$\prod_{i=1}^{n} \sqrt{\frac{\tau_e}{2\pi}} e^{-\frac{\tau_e (y_i - \mathbf{x}_i \boldsymbol{\beta})^2}{2}} = \left(\frac{\tau_e}{2\pi}\right)^{n/2} e^{-\frac{\tau_e (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2}}.$$

- The priors are

$$\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \tau_\beta, \mathbf{K}) &= \left(\frac{\tau_\beta}{2\pi}\right)^{p/2} \det(\mathbf{K})^{-1/2} e^{-\frac{\tau_\beta (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \mathbf{K}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{2}}, \\
p(\tau_\beta|\alpha_\beta, \gamma_\beta) &= \frac{\gamma_\beta^{\alpha_\beta}}{\Gamma(\alpha_\beta)} \tau_\beta^{\alpha_\beta - 1} e^{-\gamma_\beta \tau_\beta}, \\
p(\tau_e|\alpha_e, \gamma_e) &= \frac{\gamma_e^{\alpha_e}}{\Gamma(\alpha_e)} \tau_e^{\alpha_e - 1} e^{-\gamma_e \tau_e}.
\end{aligned}$$

Joint distribution

- The joint distribution
$p(\mathbf{y}, \boldsymbol{\beta}, \tau_e, \mathbf{X}, \tau_\beta) = p(\mathbf{y}|\boldsymbol{\beta}, \tau_e, \mathbf{X})p(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \tau_\beta, \mathbf{K})p(\tau_\beta|\alpha_\beta, \gamma_\beta) \times\ p(\tau_e|\alpha_e, \gamma_e)$ is

$$\left(\frac{\tau_e}{2\pi}\right)^{n/2} e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}} \times \left(\frac{\tau_\beta}{2\pi}\right)^{p/2} \det(\mathbf{K})^{-1/2} e^{-\frac{\tau_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}} \times \frac{\gamma_\beta^{\alpha_\beta}}{\Gamma(\alpha_\beta)} \tau_\beta^{\alpha_\beta - 1} e^{-\gamma_\beta \tau_\beta} \times \frac{\gamma_e^{\alpha_e}}{\Gamma(\alpha_e)} \tau_e^{\alpha_e - 1} e^{-\gamma_e \tau_e}.$$

- As the posterior distribution is proportional to the joint distribution, the task of determining conditional posteriors is equivalent to determining the distribution kernel for the parameter(s) of interest.

Constructing the Gibbs sampler

- The component of the joint distribution that is a function of τ_e is,

$$\left(\frac{\tau_e}{2\pi}\right)^{n/2} e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}} \frac{\gamma_e^{\alpha_e}}{\Gamma(\alpha_e)} \tau_e^{\alpha_e-1} e^{-\gamma_e\tau_e} \propto \tau_e^{\alpha_e+n/2-1} e^{-\tau_e(\gamma_e+(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2)}.$$

  which corresponds to a gamma kernel, therefore

$$p(\tau_e|\mathbf{y},\boldsymbol{\beta},\mathbf{X}) = \text{Ga}(\alpha_e + n/2, \gamma_e + (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2).$$

- The component of the joint distribution that is a function of τ_β is,

$$\left(\frac{\tau_\beta}{2\pi}\right)^{p/2} e^{-\frac{\tau_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}} \times \frac{\gamma_\beta^{\alpha_\beta}}{\Gamma(\alpha_\beta)} \tau_\beta^{\alpha_\beta-1} e^{-\gamma_\beta\tau_\beta} \propto \tau_\beta^{\alpha_\beta+p/2-1} e^{-\tau_\beta(\gamma_\beta+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)/2)}$$

  which corresponds to a gamma kernel, therefore

$$p(\tau_\beta|\mathbf{y},\boldsymbol{\beta},\mathbf{K}) = \text{Ga}(\alpha_\beta + p/2, \gamma_\beta + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)/2).$$

Constructing the Gibbs sampler

- The component of the joint distribution that is a function of $\boldsymbol{\beta}$ is,

$$e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}} e^{-\frac{\tau_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}} \propto e^{-\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})\boldsymbol{\beta}}{2}} e^{\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)}{2}} e^{\frac{(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)'\boldsymbol{\beta}}{2}}$$

$$= e^{-\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})\boldsymbol{\beta}}{2}} e^{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})^{-1}(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)}$$

  which corresponds to a normal kernel, hence $p(\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\boldsymbol{\beta}_0,\mathbf{K},\tau_e,\tau_\beta)$ is multivariate-normal with mean $= (\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})^{-1}(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)$ and variance-covariance matrix $(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})^{-1}$.

- The Gibbs sampler would then operate by cycling through the conditional posteriors of τ_e, τ_β and $\boldsymbol{\beta}$.

  Note: If $\mathbf{x}$ is multivariate normal $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, the kernel is $e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}} \propto e^{-\frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}{2}} e^{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mu}$

Modifying $p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}_0, \sigma_\beta^2 \mathbf{K})$ to $p(\boldsymbol{\beta}) = \mathcal{N}(\mathbf{0}_p, \sigma_\beta^2 \mathbf{K})$

- As said earlier, the prior we have just used is fairly general. Now we will consider special cases.
- Let's make the common random effect regression assumption that $\boldsymbol{\beta}_0 = \mathbf{0}$. This assumption modifies the conditional posteriors of τ_β and $\boldsymbol{\beta}$ as follows:

$$\begin{aligned} p(\tau_\beta|\cdot) &= \text{Ga}(\alpha_\beta + p/2, \gamma_\beta + \boldsymbol{\beta}'\mathbf{K}^{-1}\boldsymbol{\beta}/2) \\ p(\boldsymbol{\beta}|\cdot) &= \mathcal{N}(\tau_e(\tau_e\mathbf{X}'\mathbf{X} + \tau_\beta\mathbf{K}^{-1})^{-1}\mathbf{X}'\mathbf{y}, (\tau_e\mathbf{X}'\mathbf{X} + \tau_\beta\mathbf{K}^{-1})^{-1}) \end{aligned}$$

Modifying $p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}_0, \sigma_\beta^2 \mathbf{K})$ to $p(\boldsymbol{\beta}) = \mathcal{N}(\mathbf{0}_p, \sigma_\beta^2 \mathbf{I}_p)$

- If we go further and assume $\mathbf{K} = \mathbf{I}_p$ (again fairly common in random effect regression), then the conditional posteriors of τ_β and $\boldsymbol{\beta}$ simplify to,

$$\begin{aligned} p(\tau_\beta|\cdot) &= \text{Ga}(\alpha_\beta + p/2, \gamma_\beta + \boldsymbol{\beta}'\boldsymbol{\beta}/2) \\ p(\boldsymbol{\beta}|\cdot) &= \mathcal{N}(\tau_e(\tau_e \mathbf{X}'\mathbf{X} + \tau_\beta \mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}, (\tau_e \mathbf{X}'\mathbf{X} + \tau_\beta \mathbf{I}_p)^{-1}) \\ &= \mathcal{N}\left(\left(\mathbf{X}'\mathbf{X} + \frac{\tau_\beta}{\tau_e}\mathbf{I}_p\right)^{-1}\mathbf{X}'\mathbf{y}, \left(\mathbf{X}'\mathbf{X} + \frac{\tau_\beta}{\tau_e}\mathbf{I}_p\right)^{-1}/\tau_e\right) \end{aligned}$$

- Question: Looking at the conditional posterior of $\boldsymbol{\beta}$ when $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p/\tau_\beta)$ *a priori*, what does it remind you of.
  - Ridge regression. But how are you told normally to estimate $\lambda = \frac{\tau_\beta}{\tau_e}$?
    - Cross-validation. So we could get a situation where apart from τ_β, all other parameters are estimated in a Bayesian way.

Modifying $p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}_0, \sigma_\beta^2 \mathbf{K})$ to $p(\boldsymbol{\beta}_1) \propto 1, p(\boldsymbol{\beta}_2) = \mathcal{N}(\mathbf{0}_{p_2}, \sigma_\beta^2 \mathbf{K}_2)$

- Now consider the special case of the linear mixed model where $\boldsymbol{\beta}$ can be split into $\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ such that *a priori* β_1 and β_2 are independent, that is $p(\boldsymbol{\beta}_1) \propto 1$, $p(\boldsymbol{\beta}_2) = \mathcal{N}(\mathbf{0}_{p_2}, \sigma_\beta^2 \mathbf{K}_2)$.

- If we look at the kernel of the prior of $\boldsymbol{\beta}_2$, $e^{-\tau_\beta \boldsymbol{\beta}_2' \mathbf{K}_2^{-1} \boldsymbol{\beta}_2 / 2}$. This can be extended to incorporate the prior for $\boldsymbol{\beta}$ by assuming

$$p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto e^{-\frac{\tau_\beta \begin{pmatrix} \boldsymbol{\beta}_1' & \boldsymbol{\beta}_2' \end{pmatrix} \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{K}_2^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}}{2}} = e^{-\tau_\beta \boldsymbol{\beta}' \mathbf{K}^{-1} \boldsymbol{\beta} / 2},$$

where $\mathbf{K}^{-1} = \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{K}_2^{-1} \end{pmatrix}$.

Modifying $p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2_\beta \mathbf{K})$ to $p(\boldsymbol{\beta}_1) \propto 1, p(\boldsymbol{\beta}_2) = \mathcal{N}(\mathbf{0}_{p_2}, \sigma^2_\beta \mathbf{K}_2)$

- As a result, the conditional posteriors are modified as follows when fitting a linear mixed model,

$$
\begin{aligned}
p(\tau_\beta|\cdot) &= \text{Ga}(\alpha_\beta + p_2/2, \gamma_\beta + \boldsymbol{\beta}_2' \mathbf{K}_2^{-1} \boldsymbol{\beta}_2/2) \\
p\left(\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \middle| \cdot \right) &= \mathcal{N}\left(\tau_e \begin{pmatrix} \tau_e \mathbf{X}_1'\mathbf{X}_1 & \tau_e \mathbf{X}_1'\mathbf{X}_2 \\ \tau_e \mathbf{X}_2'\mathbf{X}_1 & \tau_e \mathbf{X}_2'\mathbf{X}_2 + \tau_\beta \mathbf{K}_2^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}, \begin{pmatrix} \tau_e \mathbf{X}_1'\mathbf{X}_1 & \tau_e \mathbf{X}_1'\mathbf{X}_2 \\ \tau_e \mathbf{X}_2'\mathbf{X}_1 & \tau_e \mathbf{X}_2'\mathbf{X}_2 + \tau_\beta \mathbf{K}_2^{-1} \end{pmatrix}^{-1} \right)
\end{aligned}
$$

- If you want to partially unblock the Gibbs sampler, and estimate $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in separate steps, the conditional posteriors are

$$
\begin{aligned}
p(\boldsymbol{\beta}_1|\cdot) &= \mathcal{N}((\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{y} - \mathbf{X}_2\boldsymbol{\beta}_2), (\mathbf{X}_1'\mathbf{X}_1)^{-1}/\tau_e) \\
p(\boldsymbol{\beta}_2|\cdot) &= \mathcal{N}(\tau_e(\tau_e \mathbf{X}_2'\mathbf{X}_2 + \tau_\beta \mathbf{K}_2^{-1})^{-1}\mathbf{X}_2'(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1), (\tau_e \mathbf{X}_2'\mathbf{X}_2 + \tau_\beta \mathbf{K}_2^{-1})^{-1})
\end{aligned}
$$

Comments

- It must be noted that the previous slides are clearly not a full summary of all possible regression type models with normal priors on coefficients.

- Some of the modifications we have not discussed include,
  - How to estimate $\mathbf{K}$.
  - Further partitioning of $\boldsymbol{\beta}$ when considering priors, for example $p(\boldsymbol{\beta}_j) \sim \mathcal{N}(\mathbf{0}_{p_j}, \sigma^2_{\beta_j}\mathbf{K}_j)$ where $\sum_{j=1}^{K} p_j = p$
  - The computational tricks that could be used to speed up estimation.

- Lastly, the assumption of a normal prior for some, but not all coefficients in the linear mixed model makes the problem look very similar to a generalised least squares problem,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

Bayesian LASSO, $p(\beta_j) \sim \text{Laplace}(\gamma)$

- In the end, we will discuss one more method of penalised regression, the LASSO (least absolute shrinkage (L1 norm) and selection operator).
- For those in the class who have used penalised regression techniques before, you may be familiar that whereas ridge regression shrinks all coefficients (L2 norm), LASSO can both shrink coefficients and set some coefficients to zero.
- From the Bayesian perspective, LASSO is equivalent to assuming *i.i.d.* Laplace priors for the regression coefficients,

$$p(\beta_j) = \frac{\gamma}{2} e^{-\gamma|\beta_j|}$$

- While you are not expected to have dealt much with the Laplace prior previously, there is one thing about the distribution you need to consider.

Bayesian LASSO, $p(\boldsymbol{\beta}_j) \sim \text{Laplace}(\gamma)$

- The Laplace distribution can be viewed as a compound distribution, just like the negative binomial or beta-binomial. For the Laplace distribution, the compound is such that $X|Y$ is normal and Y is exponential. In our particular case, X would be $\boldsymbol{\beta}_j$, and Y will be σ_j^2 as shown below,

$$\frac{\gamma}{2} e^{-\gamma|\boldsymbol{\beta}_j|} = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{\beta_j^2}{2\sigma_j^2}} \frac{\gamma^2}{2} e^{-\frac{\gamma^2\sigma_j^2}{2}} d\sigma_j^2$$

- Hence the prior of $\boldsymbol{\beta}_j$ can be extended in the following hierarchy,

$$\begin{aligned} p(\boldsymbol{\beta}_j|\sigma_j^2) &= \mathcal{N}(0, \sigma_j^2) \\ p(\sigma_j^2|\gamma) &= \text{Exp}(\gamma^2/2). \end{aligned}$$

Gibbs sampler for Bayesian LASSO

- Based on the prior specification, we see that $p(\boldsymbol{\beta}|\sigma_1^2, \ldots, \sigma_p^2)$ is an example of $\mathcal{N}(\mathbf{0}_p, \mathbf{K})$, with $\mathbf{K}$ being a diagonal matrix with $\mathbf{K}_{jj} = \sigma_j^2$. Thus we already know the conditional posterior for $\boldsymbol{\beta}$,

$$p(\boldsymbol{\beta}|\cdot) \quad = \quad \mathcal{N}(\tau_e(\tau_e \mathbf{X}'\mathbf{X} + \mathbf{K}^{-1})^{-1}\mathbf{X}'\mathbf{y}, (\tau_e \mathbf{X}'\mathbf{X} + \mathbf{K}^{-1})^{-1}).$$

- In addition, as we have not tied the prior for σ_j^2 to σ_e^2, we also already know the conditional posterior for $\tau_e = (\sigma_e^2)^{-1}$,

$$p(\tau_e|\mathbf{y}, \boldsymbol{\beta}, \mathbf{X}) = \text{Ga}(\alpha_e + n/2, \gamma_e + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2).$$

- This just leaves us needing to find the conditional posterior for σ_j^2.

Gibbs sampler for Bayesian LASSO

- You are not expected to know this distribution, but the conditional posterior for $1/\sigma_j^2$ is inverse-Gaussian with parameters $\lambda = \gamma^2$ and $\mu = \gamma/|\beta_j|$.
- Note despite the name 'inverse Gaussian' distribution, it is not the distribution of an inverse of a normal random variable. The density of an inverse Gaussian is,

$$f(x|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}},$$

with distribution kernel $x^{-3/2} e^{-\frac{\lambda x}{2\mu^2}} e^{-\frac{\lambda}{2x}}$.

Gibbs sampler for Bayesian LASSO

- To understand how this is the conditional posterior, consider the following.
- The component of the joint distribution that is a function of σ_j^2 is

$$(\sigma_j^2)^{-1/2} e^{-\beta_j^2/2\sigma_j^2} e^{-\gamma^2\sigma_j^2/2} d\sigma_j^2,$$

  which, in the parametrisation $\tau_j = (\sigma_j^2)^{-1}$ and after accounting for variable transformation in the density is,

$$(\tau_j)^{1/2} e^{-\frac{\beta_j^2\tau_j}{2}} e^{-\frac{\gamma^2}{2\tau_j}} \times \frac{1}{\tau_j^2} d\tau_j = \tau_j^{-3/2} e^{-\frac{\beta_j^2\tau_j}{2}} e^{-\frac{\gamma^2}{2\tau_j}} d\tau_j,$$

- But is there any parameter that still needs to be estimated?
  - γ. Often γ is estimated using cross-validation, or is pre-fixed. However you could attempt assigning a prior distribution to γ as well.