A funny twist on geostatistics

Ben Shaby

Department of Statistics

July 13, 2007

Generalized Poisson regression

Let $Y_i, \ldots, Y_n$ be observed counts and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ be a matrix of covariates associated with each observation. Also, let $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]$ and $\mathbf{t} = (t_1, \ldots, t_n)'$ be the locations associated with $Y_i, \ldots, Y_n$, indexed, say, by latitude, longitude, and time.

$$
\begin{aligned}
y_i | \lambda_i &\sim \text{Pois}(\lambda_i) \\
\log \lambda_i(\mathbf{x}_i; \mathbf{s}_i, t_i) &= \mu_i(\mathbf{x}_i) + u_i(\mathbf{s}_i, t_i) + \varepsilon_i \\
&\equiv b_i \\
\mathbf{u}(\mathbf{S}, \mathbf{t}) | \boldsymbol{\theta} &\sim N(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}; \mathbf{S}, \mathbf{t})) \\
\varepsilon_i | \eta &\sim \text{iid } N(0, \eta) \qquad (1)
\end{aligned}
$$

Another parametrization

We marginalize over the ε's and the u's because it will be more convenient for computation

$$\begin{aligned} y_i|\lambda_i &\sim \mathrm{Pois}(\lambda_i) \\ \log \lambda_i(\mathbf{x}_i; \mathbf{s}_i, t_i) &\equiv b_i \\ \mathbf{b}(\mathbf{X}; \mathbf{S}, \mathbf{t})|\boldsymbol{\theta}, \boldsymbol{\mu}, \eta &\sim N(\boldsymbol{\mu}(\mathbf{X}), \boldsymbol{\Sigma}(\boldsymbol{\theta}; \mathbf{S}; \mathbf{t}) + \eta\mathbf{I}) \end{aligned} \tag{2}$$

The parameters

So we have to update

- The fixed effects $\boldsymbol{\mu}(\mathbf{X})$
- The random effects, contained in the log means $\mathbf{b}$
- The covariance parameters $\boldsymbol{\theta}$ and η

None of which are straightforward.

The ridiculous Gibbs sampler

What would a Gibbs sampler look like?

$$[b_1|\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b}_{i\neq 1}]$$

$$[b_2|\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b}_{i\neq 2}]$$

$$\vdots$$

$$[b_{\text{gazillion}}|\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{b}_{i\neq \text{gazillion}}]$$

$$[\mu_1|\mathbf{y}, \mathbf{b}, \boldsymbol{\theta}, \mu_{i\neq 1}]$$

$$\vdots$$

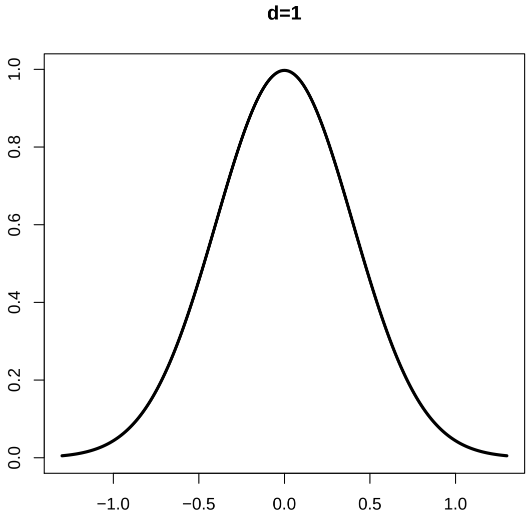
$$[\mu_{\text{gazillion}}|\mathbf{y}, \mathbf{b}, \boldsymbol{\theta}, \mu_{i\neq \text{gazillion}}]$$

$$[\alpha|\mathbf{y}, \boldsymbol{\mu}, \mathbf{b}, \sigma^2, \eta]$$

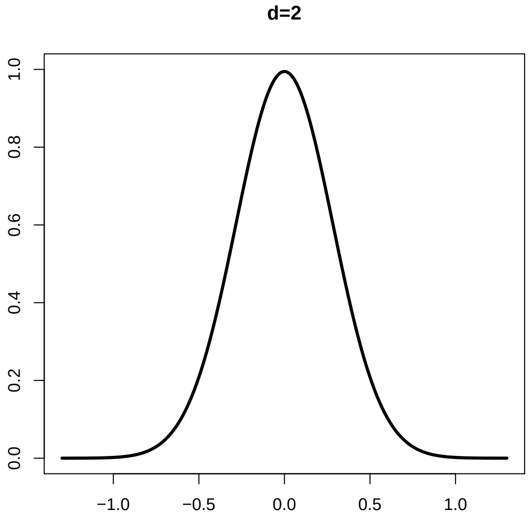
$$[\sigma^2|\mathbf{y}, \boldsymbol{\mu}, \mathbf{b}, \alpha, \eta]$$

$$[\eta|\mathbf{y}, \boldsymbol{\mu}, \mathbf{b}, \alpha, \sigma^2]$$

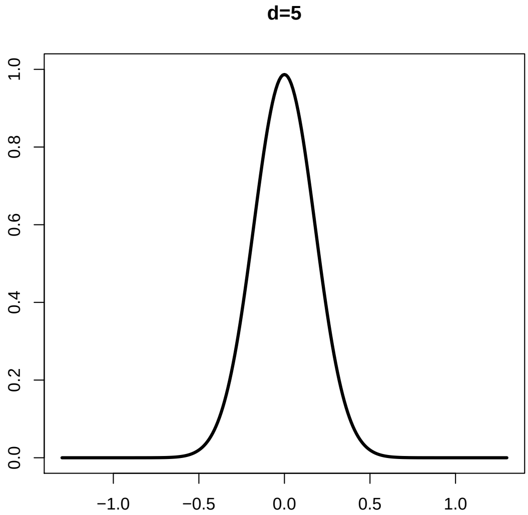
Geometric whammy



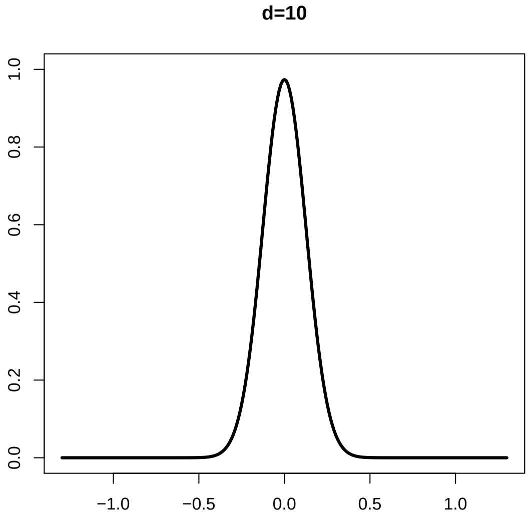
Geometric whammy



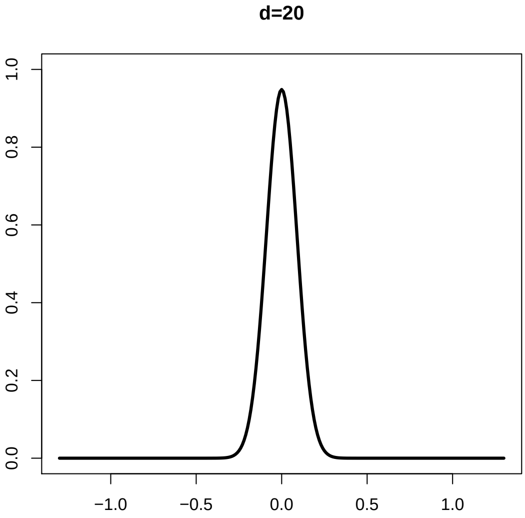
Geometric whammy



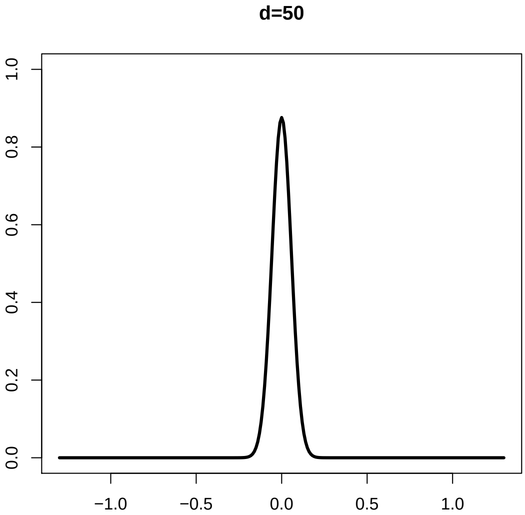
Geometric whammy



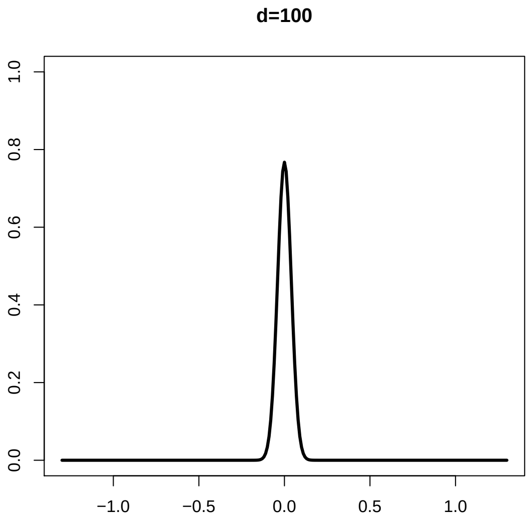
Geometric whammy



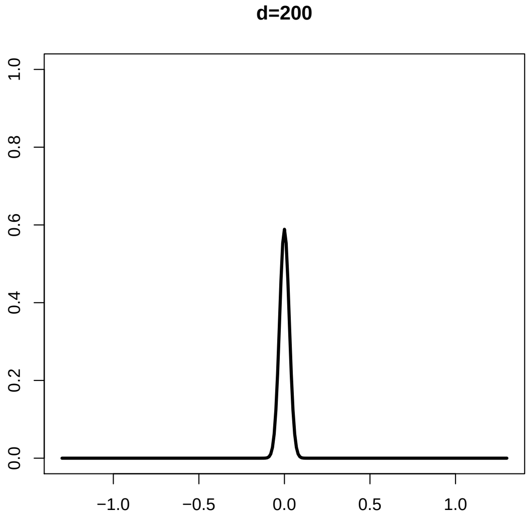
Geometric whammy



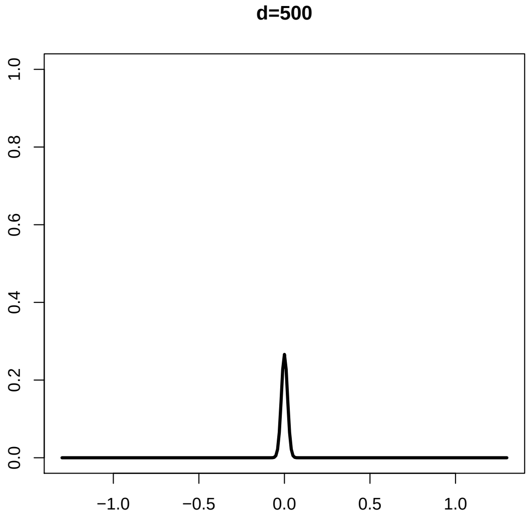
Geometric whammy



Geometric whammy



Geometric whammy



Approximating the log likelihood

- Problem: we can't compute the log likelihood function
- Solution 1: Compute many "smaller" log likelihoods and add them together
- Solution 2: Introduce zeros into the covariance matrix and use sparse matrix "magic"

Approximating the log likelihood

- Problem: we can't compute the log likelihood function
- Solution 1: Compute many "smaller" log likelihoods and add them together
- Solution 2: Introduce zeros into the covariance matrix and use sparse matrix "magic"

Approximating the log likelihood

- Problem: we can't compute the log likelihood function
- Solution 1: Compute many "smaller" log likelihoods and add them together
- Solution 2: Introduce zeros into the covariance matrix and use sparse matrix "magic"

What I think is cool about what we're doing

- We can can use data mining models inside an MCMC sampler
- Our MCMC sampler tunes itself
- We have a bunch of tricks to deal with high-dimensional data all working together

What I think is cool about what we're doing

- We can can use data mining models inside an MCMC sampler
- Our MCMC sampler tunes itself
- We have a bunch of tricks to deal with high-dimensional data all working together

What I think is cool about what we're doing

- We can can use data mining models inside an MCMC sampler
- Our MCMC sampler tunes itself
- We have a bunch of tricks to deal with high-dimensional data

all working together