Baseball Research Log

Chris Comiskey

Spring 2017

20 February 2017

********** Begin INLA-SPDE-RINLA **********

Debashis meeting

- INLA is oversold; Does Matern $\nu = 1, 2, 3$ only (no exponential, $\nu = 1/2$)
- Matern $\nu = 1$ is reasonable-ish.
- Could add structure to cov. matrix, make it sparse, make cov = 0 for $d > d_0$
- SPDE is oversell too.
- INLA is basically Taylor series
- Will be bias in estimates, for binary data in particular.
- MCMC might work better with sparse cov matrix, R has a package for sparse matrices; does Stan use sparse matrices?

February 27, 2017

Matern Covariance (1)

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu \left(\sqrt{2\nu}\frac{d}{\rho}\right)$$

Where Γ is the gamma function, and K_ν is the modified Bessel function, ρ and ν are non-negative covariance parameters

- Recall, $\Gamma(n) = (n-1)!$
- $\nu = 1/2$ gives exponential covariance

$$C_{1/2}(d) = \sigma^2 exp(-d/\rho)$$

- Stationary and isotropic if Euclidean distance

Miscellaneous

- Covariance tapering? $d < d_0 \rightarrow d = 0$, then what?
- **range parameter: the Euclidean distance where $x(s_0)$ and $x(s_1)$ are almost independent.** [Lindgren et al., 2011]

Meeting

- I **CAN** make **half** of my talk "Lessons Learned"!!
- "Lessons Learned"
  - Zero-th problem (Journal club)
  - JSM talk, already published – do a lit review!
  - Write and show
  - "Secrets of Research Success" -Hugh Kearns
  - Imposter syndrome
  - Compare yourself to others → ruin!
  - Koutsoyiannis
  - Richard Feynman
  - "Thinking, Fast and Slow"
  - "Set a pace you can keep."
- INLA
- "You're ready to be done." -Alix, as in "You have gained the research skills you need to go out there and be successful."
- "This is great! You're coming in here telling **me** what your dissertation is!"
- My plan for chapters is legit.
- Talked about family, job search, research, seminar talk, Chris Wolf.

March 6, 2017

An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach [Lindgren et al., 2011]

- "A GMRF is a discretely indexed Gaussian field x, where the full conditionals" depend only on a small set of symmetric neighbor relationships, which yields sparse matrices that lend themselves to approximations—INLA
- The GMRF "computational gain comes from the fact that the zero pattern of the precision matrix Q (the inverse covariance matrix) relates directly to the notion of neighbours..."
- Matrix **Q: the inverse covariance matrix**
- "The result is a basis function representation with piecewise linear basis functions, and Gaussian weights with Markov dependencies determined by a general triangulation of the domain."
  - Recall "basis" from linear algebra, where a set of linearly independent vectors span a space
  - function space - space of *functions*
  - basis functions - a set of functions from which can build any function in the function space
- Matern (2):

$$r(\boldsymbol{u},\boldsymbol{v}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{u}-\boldsymbol{v}||)^{\nu}K_{\nu}(\kappa||\boldsymbol{u}-\boldsymbol{v}||)$$

- scaling parameter $\kappa \approx 1/\rho = 1/\text{range parameter}$
- Empirically derived: $\rho \equiv \sqrt{8\nu}/\kappa$

- (identity) Linear fractional SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u})$$

  - $\alpha = \nu + d/2$
  - Note: d=2 for $\mathbb{R}^2$

- Δ is Laplacian:

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$$

- Marginal Variance:

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}$$

Bayesian Spatial Modelling with R-INLA [Lindgren and Rue, 2015]

- "...as discussed in Lindgren et al. (2011), one can express a large class of random field models as solutions to continuous domain stochastic partial differential equations (SPDEs), and write down explicit links between the parameters of each SPDE and the elements of precision matrices for weights in a discrete basis function representation."

- "**An alternative to traditional covariance based modelling is to use SPDEs, but carry out the practical computations using Gaussian Markov random field (GMRF) representations. This is done by approximating the full set of spatial random functions with weighted sums of simple basis functions, which allows us to hold on to the continuous interpretation of space, while the computational algorithms only see discrete structures with Markov properties.** Beyond the main paper Lindgren et al. (2011), this is further discussed by Simpson, Lindgren, and Rue (2012a,b)." [Simpson et al., 2012a], [Simpson et al., 2012b]

- Simpson et al. [2012a] give a **fantastic picture** for conveying the action of SPDE method

- "The simplest model for (spatial field) $x(\boldsymbol{s})$ currently implemented in R-INLA is the SPDE/GMRF version of the stationary Matern family, obtained as the stationary solutions to

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}),\ \boldsymbol{s} \in \Omega$$

  where

  - Δ is the Laplacian
  - κ is the spatial scale parameter
  - α controls the smoothness of the realisations
  - τ controls the variance
  - Ω is the spatial domain
  - $\mathcal{W}(\boldsymbol{s})$ is a Gaussian spatial white noise process

Whittle (1954, 1963) shows stationary solutions on $\mathbb{R}^d$ have Matern covariances,

$$\text{COV}(x(\mathbf{0}), x(\boldsymbol{s})) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||s||)^\nu K_\nu(\kappa||s||)$$

The parameters in the two formulations are coupled so that the Matern smoothness is $\nu = \alpha - d/2$ and marginal variance is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}$$

Exponential covariance: $\nu = 1/2$; (i) for $d = 1 \to \alpha = 1$, (ii) for $d = 2 \to \alpha = 3/2$"

- "The models discussed in Lindgren et al. [2011] and implemented in R-INLA are built on a basis representation

$$x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k$$

where

  - $\psi_k(\cdot)$ are deterministic basis functions, and
  - the joint distribution of the weight vector $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is chosen so that the distribution of the functions $x(\boldsymbol{s})$ approximates the distribution of solutions to the SPDE on the domain."
  - piecewise polynomial basis functions
  - use **Finite Element Method - project the SPDE onto the basis representation** ...which is GMRF. [Simpson et al., 2012a]

- Bayesian Inference (REALLY FREAKIN' GOOD SUMMARY)

  1. $\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ by Gaussian approximation
  2. Posterior mode: $\boldsymbol{x}^*(\boldsymbol{\theta}) = \text{argmax}_x p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$
  3. Laplace approximation

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{p(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x})}{p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})} \approx \frac{p(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x})}{\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}$$

     Approximate (unnormalized) posterior density for $\boldsymbol{\theta}$ at any point, and numerical optimization to find mode of posterior.

  4. Numerical integration:

$$p(\theta_k|\boldsymbol{y}) \approx \int \tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-k}$$

  5. Numerical integration:

$$p(x_j|\boldsymbol{y}) \approx \int \tilde{p}(x_j|\boldsymbol{\theta}, \boldsymbol{y}) \tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}$$

Miscellaneous

I got `inla(...)` to run!!

- `geoR` package
- I am trying to generate Matern data with `grf(...)`, then estimate parameters with `inla(...)`. However, basic input to `inla(...)` implies INLA, without SPDE connection, and thus data on a grid. This means using my actual pitch locations from `hitter` will not work.

- Try to generate data on a grid, use `inla(...)` to estimate parameters
- inverse Precision matrix equals covariance matrix: $Q^{-1} = \Sigma$
- Precision matrix Q equals inverse covariance matrix: $Q = \Sigma^{-1}$

Simulate Matern data

`geoR` package, `grf(...)` function

Matern (3)

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

- u - vector/matrix/array with distances between pairs of data locations
- ϕ - range parameter, > 0
- κ - smoothness parameter, > 0
- $K_{\kappa}(\cdot)$ - modified Bessel function of third kind of order κ

Matern (4)

[Schabenberger and Gotway, 2004]

$$C(h) = \sigma^2 \frac{1}{\Gamma(\nu)} \left(\frac{\theta h}{2}\right)^{\nu} 2K_{\nu}(\theta h),\ \nu > 0, \theta > 0$$

- θ governs range of spatial dependence
- ν: smoothness increases with ν
- Page 143: The Matern Class of Covariance Functions
- Page 199: **THREE** Matern parameterizations
- Page 210: Bessel Functions
- Whittle Model, $\nu = 1$ (INLA/SPDE world, I think)

  $$C(h) = \sigma^2 \theta h K_1(\theta h)$$

  Whittle considered this the "elementary model" in $\mathbb{R}^2$

Stochastic Calculus

Stochastic Differential Equation (wikipedia)

- "A heuristic but helpful interpretation of the stochastic differential equation (of continuous time stochastic process X_t) is that in a small time interval of length δ the stochastic process X_t changes its value by an amount that is normally distributed (for example) with expectation $\mu(X_t, t)\delta$ and variance $\sigma(X_t, t)^2\delta$ and is independent of the past behavior of the process."

Stochastic Calculus [Mao, 2007]

- Integral of random process is another random process. Random process not integrable in traditional sense, so stochastic calculus created, by Ito, Ito Calculus. Use Brownian motion as some sort of reference point.

  $$Y_t = \int_0^t H_s dX_s$$

  Integrand and integrator are stochastic processes.

March 16, 2017

- Smoothness = differentiability, number of continuous derivatives (Wikipedia)
- Stochastic calculus = "branch of mathematics that operates on stochastic processes. It allows a consistent theory of integration to be defined for integrals of stochastic processes with respect to stochastic processes. It is used to model systems that behave randomly.

  The best-known stochastic process to which stochastic calculus is applied is the Wiener process (named in honor of Norbert Wiener), which is used for modeling Brownian motion as described by Louis Bachelier in 1900 and by Albert Einstein in 1905 and other physical diffusion processes in space of particles subject to random forces. Since the 1970s, the Wiener process has been widely applied in financial mathematics and economics to model the evolution in time of stock prices and bond interest rates. The main flavours of stochastic calculus are the Ito calculus and its variational relative the Malliavin calculus." (Wikipedia)
- Finite element method - "a numerical method for solving problems of engineering and mathematical physics... The finite element method formulation of the problem results in a system of algebraic equations. The **method yields approximate values of the unknowns at discrete number of points over the domain. To solve the problem, it subdivides a large problem into smaller, simpler parts that are called finite elements.** The simple equations that model these finite elements are then assembled into a larger system of equations that models the entire problem. FEM then uses variational methods from the calculus of variations to approximate a solution by minimizing an associated error function." (Wikipedia)

MEETING

- Graduate this summer. Pay for three credits out of pocket.
- Need to email committee, see when members are available.
- Need to investigate the term options, see which term I will enroll for.
- Need to procure insurance for the summer.
  - Need to talk to insurance office about this
- April 24 Seminar: Alix is reneging on "Lessons Learned" portion
  - Jeffrey really impressed her with his talk, and thus Lan impressed her.
  - ...so she wants to show me off, in a sense.
  - I gotta keep doing what I did to get to the majors.
  - ...Don't do anything different. Stick with your approach.
- Alix really thinks "so much of life is just showing up."
- Fall GTA is off the table.
- Alix wants me to come up with a timeline of when I will give she and Charlotte my Chapter 1, Chapter 2, Chapter 3 drafts; I agreed to middle of Spring term to provide this timeline. (deadline for the timeline)

March 27, 2017

Lecture Slides - GMRF, Dependent Spatial Data

[Lindstrom et al., 2011]

- Define problem, explain difficulty, define GMRF, **sparse** precision matrix Q^{-1}, Markov property, Precision matrix construction hard
- Matern family:
$$r(\boldsymbol{u}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{u}||)^\nu K_\nu(\kappa||\boldsymbol{u}||)$$
- Random fields with Matern covariance are solutions to:
$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}),$$
  - $\alpha = \nu + d/2$
  - Δ is Laplacian: $\Delta = \sum_i \frac{\partial^2}{\partial x_i^2}$

Construct GMRF from SPDE

(by representing Matern field as a basis function (which is GMRF), using the SPDE identity, in what amounts to finite element method (FEM))

- Construct the solution as a finite basis expansion:
$$x(\boldsymbol{s}) = \sum_k \psi_k(\boldsymbol{s}) x_k,$$
with a suitable distribution for the weights $\{x_k\}$.
- Sotchastic weak solution given by weights $\{x_k\}$ such that the joint distribution fulfills
$$\sum_i \left\langle \psi_j, (\kappa^2 - \Delta)^{\alpha/2} \psi_i x_i \right\rangle \stackrel{D}{=} \langle \psi_j, \mathcal{W} \rangle \ \forall \boldsymbol{j}$$
- SOLUTION.
The distribution of the weights:
$$\boldsymbol{x} \sim N(0, \boldsymbol{Q}^{-1}) \tag{1}$$
$$\alpha = 1 : \boldsymbol{Q}_{1,\kappa} = \boldsymbol{K} \tag{2}$$
$$\alpha = 2 : \boldsymbol{Q}_{2,\kappa} = \boldsymbol{K}\boldsymbol{C}^{-1}\boldsymbol{K} \tag{3}$$
$$\alpha = 3, 4, \ldots : \boldsymbol{Q}_{\alpha,\kappa} = \boldsymbol{K}\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2,\kappa}\boldsymbol{C}^{-1}\boldsymbol{K}, \tag{4}$$
where
$$\boldsymbol{C}_{i,j} = \langle \psi_i, \psi_j \rangle \tag{5}$$
$$\boldsymbol{K}_{i,j} = \langle \psi_i, (\kappa^2 - \Delta)\psi_j \rangle \tag{6}$$
$$= \kappa^2 \langle \psi_i, \psi_j \rangle - \langle \psi_i, \Delta \psi_j \rangle \tag{7}$$
$$= \kappa^2 \boldsymbol{C}_{i,j} + \boldsymbol{G}_{i,j} \tag{8}$$
- $\boldsymbol{C}^{-1}$ dense, so replace $\boldsymbol{C}$ with diagonal matrix $\widetilde{\boldsymbol{C}}_{i,i} = \langle \psi_i, \boldsymbol{1} \rangle$

Example: Lattice on $\mathbb{R}^2$, step size h, regular triangulation

- $\boldsymbol{Q}_{2,\kappa} = \kappa^4 h^2 + 2\kappa^2(-\Delta) + \frac{1}{h^2}\Delta^2$
- $\boldsymbol{Q}_{2,\kappa} \approx \kappa^4 h^2 M_0 + 2\kappa^2 M_1 + \frac{1}{h^2} M_2$
- $M_0 = \begin{bmatrix} & & & & \\ & & 1 & & \\ & & & & \\ & & & & \end{bmatrix}$
- $-\Delta \approx M_1 = \begin{bmatrix} & & & & \\ & & -1 & & \\ & -1 & 4 & -1 & \\ & & -1 & & \\ & & & & \end{bmatrix}$
- $\Delta^2 \approx M_2 = \begin{bmatrix} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{bmatrix}$
- Matrix form, matrix $\boldsymbol{A}$:
  - $A_{i\cdot} = \left[\psi_1(\boldsymbol{u}_i), \ldots, \psi_N(\boldsymbol{u}_i)\right]$
  - $x \in N(\mu, \boldsymbol{Q}^{-1})$
  - $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$

Stat Methods for Spatial Data Analysis

Matern, Whittle, Stochastic Laplace [Schabenberger and Gotway, 2004]

- Matern (θ range, smoothness ν)

$$C(h) = \sigma^2 \frac{1}{\Gamma(\nu)} \left(\frac{\theta h}{2}\right)^{\nu} 2K_\nu(\theta h),\ \nu > 0, \theta > 0$$

- Whittle Model ($\nu = 1$) (Whittle: "elementary model" in $\mathbb{R}^2$)
  Covariance, Correlation:

$$C(h) = \sigma^2 \theta h K_1(\theta h)$$
$$R(h) = \theta h K_1(\theta h)$$

- Stochastic Laplace equation (Z: Matern, ϵ: white noise):

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \theta^2\right) Z(x, y) = \epsilon(x, y)$$

Lecture Slides - GMRFs

[Lindstrom, 2014]

- Kriging
  1. Observations $Y(\boldsymbol{s}_i), i = 1, \ldots, n$, and unobserved locations $X(\boldsymbol{s})$.
  2. Simplest case, Gaussian
  $$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx}^T & \Sigma_{xx} \end{bmatrix} \right)$$
  3. Parametric form
  $$\boldsymbol{Y} \sim N(\boldsymbol{\mu}(\boldsymbol{\theta}), \Sigma(\boldsymbol{\theta}))$$
  4. Log-likelihood
  $$L(\boldsymbol{\theta}|\boldsymbol{Y}) = -\frac{1}{2} log|\Sigma(\boldsymbol{\theta})| - \frac{1}{2}\Big(\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\theta})\Big)^T \Sigma(\boldsymbol{\theta})^{-1}\Big(\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\theta})\Big)$$
  5. Krig
  $$E\Big(\boldsymbol{X}|\boldsymbol{Y}, \hat{\theta}\Big) = \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_y\Big)$$
- "Big N" problem
- Getting aroung "Big N"; lots of references
- Gaussian Markov Random Field (GMRF) - AR1 is simplest example
- Precision matrix $Q = \Sigma^{-1}$
- GMRF: neighbour structure $\rightarrow$ sparse precision matrix $\rightarrow$ simplified conditional expectation
- Brief computational details - sparse $\boldsymbol{Q}$ $\rightarrow$ sparse $\boldsymbol{R}$ (Cholesky: $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{Q}$)
- $\boldsymbol{Q}$ not trivial to construct; create $\boldsymbol{Q}$ from Matern as solution to SPDE
- Matern produces Markov field for $\nu \in \mathbb{Z}$ for $\mathbb{R}^2$ (...via non-trivial process projection of Matern, with SPDE, onto basis representation)
- Construct a discrete approximation of the continuous field using basis functions, $\{\psi_k\}$, and weights $\{w_k\}$,
$$x(\boldsymbol{s}) = \sum_k \psi_k(\boldsymbol{s}) w_k$$
- Find distribution of w_k by solving
$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$$
- **Stochastic weak solution** to the SPDE:
$$\Big[\Big\langle \phi_k, (\kappa^2 - \Delta)^{\alpha/2}\boldsymbol{x}\Big\rangle\Big]_{k=1,\ldots,n} \stackrel{D}{=} \Big[\langle \phi_k, \mathcal{W}\rangle\Big]_{k=1,\ldots,n}$$
...for each set of test functions $\{\phi_k\}$.
- Definition [Lindgren et al., 2011]:
$$\langle f, g\rangle = \int f(\boldsymbol{u}) g(\boldsymbol{u}) d\boldsymbol{u}$$

- Replace $\boldsymbol{x}$ with basis function representation $\Sigma_k \psi_k w_k$

$$\left[\left\langle \phi_i, (\kappa^2 - \Delta)^{\alpha/2} \psi_j \right\rangle\right]_{i,j} \boldsymbol{w} \stackrel{D}{=} \left[\langle \phi_k, \mathcal{W} \rangle\right]_k$$

- Galerkin solution: $\alpha = 2, \phi_i = \psi_i$

$$\left(\kappa^2[\langle \psi_i, \psi_j \rangle] + [\langle \psi_i, -\Delta \psi_j \rangle]\right) \boldsymbol{w} \stackrel{D}{=} \left[\langle \psi_k, \mathcal{W} \rangle\right]$$

$$\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right) \boldsymbol{w} \stackrel{D}{=} N(0, \boldsymbol{C})$$

$$\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right) \boldsymbol{w} \sim N(0, \boldsymbol{C})$$

Variance(w)

$$Var\left[\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right) \boldsymbol{w}\right] = \boldsymbol{C} \tag{9}$$

$$\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right) Var(\boldsymbol{w}) \left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^T = \boldsymbol{C} \tag{10}$$

$$Var(\boldsymbol{w}) = \left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^{-1} \boldsymbol{C} \left(\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^T\right)^{-1} \tag{11}$$

$$\boldsymbol{Q}^{-1} = \left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^{-1} \boldsymbol{C} \left(\left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^T\right)^{-1} \tag{12}$$

$$\boldsymbol{Q} = \left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right)^T \boldsymbol{C}^{-1} \left(\kappa^2 \boldsymbol{C} + \boldsymbol{G}\right) \tag{13}$$

- **Piecewise linear basis function gives (almost) GMRF** [Lindgren et al., 2011]. $\boldsymbol{G}_{ij}$ and $\boldsymbol{C}_{ij}$ sparse, but $\boldsymbol{C}_{ij}^{-1}$ not. Replace $\boldsymbol{C}$ with diagonal matrix $\widetilde{\boldsymbol{C}}$

$$\widetilde{\boldsymbol{C}}_{i,i} = \int \psi_i(\boldsymbol{s}) d\boldsymbol{s}$$

- Regular lattice in $\mathbb{R}^2$, order $\alpha = 2(\nu = 1)$, same M_0, M_1, M_2 as above.
- $\boldsymbol{A}$ matrix, with rows $\boldsymbol{A}_{i\cdot} = [\psi_1(\boldsymbol{s}_i), \ldots, \psi_N(\boldsymbol{s}_i)] \rightarrow \boldsymbol{x}(\boldsymbol{s}) = \boldsymbol{A}(\boldsymbol{s})\boldsymbol{w}$ where $\boldsymbol{w} \sim N(\mu, \boldsymbol{Q}^{-1})$
- Observations, kriging $E(w|y)$ (I have a logistic link between y and w, so kriging estimates not available)
- For me: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{w}$ (X = covariates, random effect Z = Aw)

Bayesian Hierarchical Model using GMRF

- (1) $p(y|\eta, \theta)$, (2) $p(\eta|\theta)$, (3) $p(\theta)$ (they use X for "latent field,", whereas I have η and Z)
- For INLA require: $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = \prod_i p(y_i|x_i, \theta)$
- Interested in:
  1. Posterior for parameters: $p(\theta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$
  2. Posterior for latent field $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) = \int p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{\theta}|\boldsymbol{y})d\theta$
- Stats 101: $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$

$$\rightarrow p(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})}$$

- (from (1) above)
$$p(\theta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} \cdot p(\boldsymbol{\theta})$$
Denominator is challenge
- Stats 101: $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$
Cut out middle man:
$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$$
Divide both sides:
$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})/p(\boldsymbol{y}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$$
y and θ constants so proportionality:
$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) \propto p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$$
Log of both sides:
$$\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) + \log p(\boldsymbol{x}|\boldsymbol{\theta}) + \text{constant}$$
- Second order Taylor approximation of $f(x) = \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ about x_0
- Obtain Gaussian approximation $p_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$:
$$E_{x_0}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \approx \ldots$$
$$V_{x_0}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \approx \ldots$$

Integrated Nested (Taylor) Laplace Approxmiation

Evaluate $p(\theta|\boldsymbol{y})$:

1. For given θ find mode: $x_0 = \text{argmax}_x p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ using $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ derived from Stats 101.
2. Compute Taylor expansion of $f(x) = \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ about x_0
3. Approximation:
$$p(\theta|\boldsymbol{y}) \approx \tilde{p}(\theta|\boldsymbol{y}) \propto \frac{p(\boldsymbol{y}|\boldsymbol{x}_0,\boldsymbol{\theta})p(\boldsymbol{x}_0|\boldsymbol{\theta})}{p_G(\boldsymbol{x}_0|\boldsymbol{y},\boldsymbol{\theta})} \cdot p(\boldsymbol{\theta})$$
4. (Approximate) MLE is: $\hat{\theta}_{\text{ML}} \approx \text{argmax}_\theta \tilde{p}(\theta|\boldsymbol{y})$

Posteriors for $[x|y]$, use numerical integration over θ:
$$p(x_i|\boldsymbol{y}) = \int p(x_i|\theta,\boldsymbol{y})p(\theta|\boldsymbol{y})d\theta \approx \sum_k p_G(\boldsymbol{x}_0|\theta_k,\boldsymbol{y})\tilde{p}(\theta_k|\boldsymbol{y})$$

Lecture Slides - Latent Gaussian Processes and SPDEs

[Lindstrom, 2016]

- More of the same. Thank god.

Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Field Models [Rue and Martino, 2007]

- $\mathcal{O}$ definition, Wikipedia

$$f(x) = \mathcal{O}(g(x)) \text{ as } x \to \inf \text{ iff } \exists M, x_0 \text{ such that } |f(x)| \leq M|g(x)| \text{ for all } x \geq x_0$$

- Gaussian Markov Random Field
  - $\boldsymbol{x} = \{x_i : i \in \mathscr{V}\}$
  - "$\boldsymbol{x}$ is a $n = |\mathscr{V}|$-dimensional Gaussian random vector with additional conditional independence/-Markov properties"
  - $\mathscr{V} = \{1, \ldots, n\}$
  - Graph $\mathscr{G} = \{\mathscr{V}, \mathscr{E}\}$, with vertices, edges.
  - "Two nodes, x_i and x_j are conditionally independent given the remaining elements of $\boldsymbol{x}$, if and only if $\{i, j\} \notin \mathscr{E}$.
  - "Then $\boldsymbol{x}$ is a GMRF with respect to $\mathscr{G}$."
  - "The edges in $\mathscr{E}$ are in one-to-one correspondence with the non-zero elements of the precision matrix of $\boldsymbol{x}$, $\boldsymbol{Q}$, in the sense that $\{i, j\} \in \mathscr{E}$ if and only if $Q_{ij} \neq 0$ for $i \neq j$."
  - "When $\{i, j\} \in \mathscr{E}$ we say that i and j are neighbours, which we denote $i \sim j$."
- "A hierarchical model:
  - **First stage**: distributional assumptions for observables conditional on latent parameters. Given observational model parameters, often assume observations conditionally independent.
  - **Second stage**: prior model for latent parameters, or (link) function of them. At this stage **GMRFs provide a flexible tool to model the dependence** between the latent parameters and thus, implicitly, the dependence between the observed data.
  - **Third stage**: prior distributions for unknown hyperparameters (precision parameters in the GMRF)."
- "We propose a deterministic alternative to MCMC based inference... computed almost instant[ly]... proves to be quite accurate."

INLA

1. $p(y|x, \theta)$ — Easy.
2. $p(x|\theta, y)$ — Gaussian approximation.
3. $p(\theta|y)$ — Identity, Gaussian approximation, evaluate at x_0
4. $p(x_i|y)$ — Numerical integration of $p(x_i|\theta, y)p(\theta|y)$ over all θ
5. $p(x|y)$ — Never.
6. $p(x, \theta|y)$ — Never.

INLA

1. $p(y|x,\theta)$ — Easy.
2. $p(x|\theta,y)$ — Gaussian approximation.

   [Rue and Held, 2005]

   - (Chapter 4) - "$\pi(x|\theta,y)$ can often be well approximated with a Gaussian distribution, by matching the mode and curvature at the mode."

   [Rue and Martino, 2007]

   - $p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_i \log p(y_i|x_i)\right)$
   - $p_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q}+\text{diag}(\boldsymbol{c}))(\boldsymbol{x}-\boldsymbol{\mu})\right)$
     - μ = mode of $p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ for each θ [Rue and Martino, 2007]
     - Terms of $\boldsymbol{c}$ due to 2nd order Taylor expansion at $\mu = x_0 =$ mode of $\sum_i \log p(y_i|x_i) = \log f(y|x)$

   [Lindstrom, 2014]

   - For a given $\theta = \theta_0$ find the mode: $x_0 = \text{argmax}_x p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}_0)$
     - Can find from unnormalized $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$, which we have up to a proportionality constant (good enough for mode)
     - $p_G(\cdot)$ is valid pdf
   - $\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) + \log p(\boldsymbol{x}|\boldsymbol{\theta}) + \text{constant}$
   - Second order Taylor approximation of $f(x) = \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ about x_0
   - Obtain Gaussian approximation $p_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$:

$$E_{x_0}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \approx \left(\boldsymbol{Q} - \text{diag}(f''(x_0))\right)^{-1}\left(\boldsymbol{Q}\boldsymbol{\mu} + f'(x_0) - f''(x_0)x_0\right) \tag{14}$$

$$V_{x_0}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \approx \left(\boldsymbol{Q} - \text{diag}(f''(x_0))\right)^{-1} \tag{15}$$

3. $p(\theta|y)$ — Summary of steps:

   (i) identity
   (ii) mode of $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}_0)$ for θ_0
   (iii) Gaussian approximation $p_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$,
   (iv) $\theta_{ML} = \text{argmax}_\theta p(\theta|y)$

   - (Ends up being same as Tierney and Kadane [1986] Laplace Approximation)
   - Stats 101 Identity: $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$
   - Bayesian 101, and identity:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} \cdot p(\boldsymbol{\theta})$$

   - **Mode** as above: For a given $\theta = \theta_0$ find the mode: $x_0 = \text{argmax}_x p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}_0)$
   - Gaussian approximation $p_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ from above.
   - Approximation of $p(\theta|\boldsymbol{y})$:

$$\tilde{p}(\theta|\boldsymbol{y}) \propto \frac{p(\boldsymbol{y}|\boldsymbol{x}_0,\boldsymbol{\theta})p(\boldsymbol{x}_0|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p_G(\boldsymbol{x}_0|\boldsymbol{y},\boldsymbol{\theta})}$$

- $\theta_{ML} = \text{argmax}_\theta \tilde{p}(\theta|y)$

4. $p(x_i|y)$ — Numerical integration of $p(x_i|\theta, y)p(\theta|y)$ over all θ
    - Posteriors for $[x|y]$, use numerical integration over θ:

$$p(x_i|\boldsymbol{y}) = \int p(x_i|\theta, \boldsymbol{y})p(\theta|\boldsymbol{y})d\theta \approx \sum_k p_G(\boldsymbol{x}_i|\theta_k, \boldsymbol{y})\tilde{p}(\theta_k|\boldsymbol{y})$$

5. $p(x|y)$ — Never.
6. $p(x, \theta|y)$ — Never.

Think continuous: Markovian Gaussian models in spatial statistics

[Simpson et al., 2012a]

- **3.4. Approximating Gaussian random fields: the finite element method)**
-

April 3, 2017

SPDE-INLA summary: A fancy approximation technique called INLA works well for Bayesian hierarchical models with latent Gaussian markov random fields, which are discrete, and have a sparse precision matrix—which is essential. To make this technique work for continuous domain spatial data and continuous spatial latent gaussian random fields, a fancy identity comprised of a linear fractional stochastic partial differential equation helpfully relates a (continuous) Matern random field to Gaussian random field white noise process. This identity undergirds approximating the Matern with a piecewise linear basis representation, with deterministic basis functions and GMRF weights. The method covers the domain with a triangular mesh to define the piecewise linear basis function representation. This process, used extensively in other fields, is known as the Finite Element Method. The basis representation has a sparse precision matrix, and enables INLA approximation techniques that offer so many computational advantages.

Work

- Points are too fucking close together. Unnecessarily close together. I can barely generate Matern data (n = 3000) with `grf(...)` because the matrix operations crash.
- Solution: use Var-res grid structure to discretize to binomial, then INLA procedure.
- If I generate Binomial(10,000, p) at all locations, where $(\sigma_0, \rho_0) = (\sigma, \rho)$, so $(\theta_1, \theta_2) = (0, 0)$, it closes in on:

```
Model hyperparameters:
                  mean     sd 0.025quant 0.5quant 0.975quant  mode
Theta1 for obs 0.1270 0.0331     0.0625   0.1268     0.1925 0.126
Theta2 for obs 0.0466 0.0783    -0.1087   0.0471     0.1997 0.049
```

  ...which I'm betting is the bias Debashis was talking about. This is **drastically, WAAAAAAAY** closer to zero that it was with Bernoulli at each location, but it won't seem to get any closer.

- In(LA) s(PDE)ummary, it is FAST but weak. Like Billy Hamilton. Maybe INLA alone, with Var-res HM box centers/binomial data would be better.

Meeting

- Graduate Committee
  1. Email Lisa about replacing her
  2. Email Sarah about availability, interest, time
  3. Find out graduate school procedure for replacing committee member
  4. Line up members availability, get okay, schedule with Grad school
- Seminar - April 24
  - Use 8 year-old Royals Chris picture to break the ice, put everyone at ease, convey lifelong passion. Then HOF pic at end?
  - ESPN clip is a go!
  - Meet with Alix the week before to go over the slides, no later than Wednesday (April 19th)
  - The story
    1. Part of zeroth problem: fast enough for TV.
    2. Variable-resolution empirical heat hap, dynamic heat map CIs
    3. HMC in Stan... too big.
    4. PPMs in `spBayes`... victory!
    5. Next steps... INLA
    6. Next steps... scoring rule evaluations
  - Tread carefully with Q & A type stuff. Alix typically finds it patrionizing. However, (her idea) I could set the stage by describing why opinions matter, and/or call on my *peers* to give opinions (rather than question open to all) — "So Joe, what do you think..." Then Alix said "I shouldn't have said anything... I don't want to squelch your style."
  - Charlotte said to check out "Exponent" consulting
  - Alix said to check out "Verry Consulting," run by her graduate school buddies from CM.
- Post-docs
  - Without publications, I'm fucked.
  - Run it by them before I take the time to apply.

April 12, 2017

- Remove `fill = Hitting` from this line:

  ```
  ggplot(...) +
  geom_text(aes(fill = Hitting, label = Count), size = 3.5)
  ```

  to eliminate warning message.

April 13, 2017 — Meeting (All seminar prep)

- Presentation looks great!
- Some changes to make...
- Sample size and resolution go hand and hand; explain. The reason we stop dividing the 22 box is because that is a small sample size.
- Need a sequi from HM talk to "Let's Model". Motivate it with "We think this is a smooth surface..."
- For my thesis, need to fix strike zone box, so that more rectangular
- Define i = ?, j = ?, on the "Let's Model" slide
- Add a bit about all the things we leave out on the "Let's Model" slide, and why; pitcher, weather, park, umpire, count, etc.
- Go ahead and show off Fleisig (ASMI), Dowling (Motus)
- Use static Shiny App to give the current "state of the art", with three boxes
- Explain at the beginning why $\hat{p}$ so low; Success vs. BA etc, when I show the first heat map
- When Bayes Hier. Model — should we emphasize challenge? Latent, unobserved, correlation...
- Segui, segui, segui; need to know what is on the next slide (bullet point, or sentence)
- Peralta the whole way through
- add p_s for location
- Hold back Big O(n^3) until second big N slide
- Give reference on first PPM slide (and Stan, for that matter)
- PPM - more discussion. Give a picture of 3000 pitch guy (if that is what I can do), tie it in to Stan's n = 300, t = 6 hours specs... "We still need to do better, need estimate for n = 9000 $\rightarrow$ INLA..."
- Need to meet earlier than normal next week, to go over slides and have time to make changes
- Explain (1,3) works because we know, but (red, blue) doesn't because...
- Explain why empirical to model is necessary

April 20, 2017 - Meeting

- Showed A & C my slides. "This is fantastic." -Alix
- Some small-ish changes.
- Sarah loved my Shiny app.
- "This is great Chris, and it was totally your idea. You should be really proud of this" -Alix, on Shiny and Var-res
- Regarding SPDE-INLA stuff on last slide of presentation: "I'm so proud of you Chris!" -Alix
- Now the team is thinking—maybe Ch. 3 is an R package or two, instead of simulation/validation study.

- Lots of talk about this, and the validity (or not) of it as a thesis component/chapter.
- Jobs. Plan B – work from Corvallis.
- Maybe I should email Stuart, see if working remote from Corvallis is viable.
- Website, labs ("You did **such** a good job on those!" -Alix), a page or two on each consulting project, baseball plots/heat maps, link to my two packages (theoretically), my publication (!)
- Charlotte said many things I have done are not that different from the things on the "Simply" girl's website, I just need to think of and package it differently.
- She's right.
- Everyone thinks I should publish my var-res and Shiny stuff. And/or make a package.
- Then put it on my website!
- Joe got a Gore offer, and an interview at UW.
- I should put a local Monster resume up.
- I should make a US resume
- Ch 3 = Package does a lot more for job opportunities than a simulation study!

April 25, 2017

- `inla(...)` is coming up with same coefficient estimates as `glm(...)`!!! It actually works!!
- Don't forget:
  - Scoring rules - model evaluation/comparison (two papers)
  - Spatial design analysis - knot selection (two papers)
  - Convergence diagnostics (CODA package, Finley paper)

Meeting - April 27, 2017

- Write, write, write. As of now, make writing my top priority.
- "While listening to your seminar, I thought to myself 'Dude's done a lot of stuff!' So, way to go, well done." -Alix
- Alix **really** wants a Shiny CI for PPM fit.
- Packages are a go!
- "Oh I got over it." -Alix, regarding her previous reservations about package building for a dissertation, when I asked about them.
- Make my personal webpage on Github. It looks better to have a slightly less impressive page, but for them to know I wrote some code in making it.
- Pay for EARL conference? Forget it.
- Seminar feedback
  - Guy asked a question about the "box", and I interpreted it as a strike zone question, A & C interpreted it as a grid box (stats!) question.

- A recurring theme was: My baseball passion occasionally occludes my statistics training. i.e. answer to "box" question, answer to Joe's question
- Comparing the effect of resolution on heat maps to the effect of bin width on histograms is a very useful comparison; would have helped my presentation.
- The var-res heat maps innovation was not clearly distinct, did not clearly stand apart—as it should. Define/present the heat map resolution selection **problem**; then provide **solution**. This will make it more clear, distinct. In my presentation it was sort of woven in. Make it clear it was a separate innovation. (I had already thought most of this bullet!)
- Spell check! Alix said she noticed multiple typos.
- Slide 46; subscript issue
- Subscript use confusing. The pitch, box, location distinctions are lost/buried in the notation. Comes out confusing, unclear.
- Think about Joe's streakiness question. Have a couple of references in my back pocket with regard to streakiness, and have a statistician's answer; not a baseball fan's answer (as I did in seminar). When they asked me in the meeting, I gave them my actual answer, which is quite sophisiticated and well thought out; that is the appropriate level of statistical professionalism. (non-stationary p!)
- They liked that people were asking questions!
- Outline slide (that I removed) would have helped. Neither of them *generally* like outline slides, but both thought it would have helped, even if I had presented the four points verbally and referred back to them.
- I may give this as a job talk, so think about other areas where this research extends, particularly within the realm of the particular job I may be applying to.
- Again–professional statistician, not baseball player. Try to somehow get to the point where that is my default take on questions (such as 'box' question).

Dissertation Chapters Outline

Priority 1: write!!

1. Introduction
2. Variable Resoluation Heat Maps
    - Chapter Appendix: R package info (Priority 3-A)
3. Shiny
    - Details of GLM, with biomechanics and Shiny demo
    - Chapter Appendix: R package info (Priority 3-B)
4. Stan, PPM, INLA
    - Shiny CI results comparison: GLM vs. PPM (**Priority 2**; Alix hyped up about this one)
    - Performace of GLM, PPM, INLA (scoring rules) (if time allows... but it better, if at all possible)
5. Conclusion (could be 5 pages)

Miscellaneous

- In the Gaussian approximation first step of INLA, we calculate a Taylor series expanded about x_0 of/for $\sum_i \log p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta})$, to build p_G. It is an exceedingly non-trivial, iterative computational algorithm (estimating x_0, and estimating $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ with p_G. In the next step we evaluate p_G at x_0; but we don't evaluate the Taylor series itself, which is now deeply subsumed in p_G.
-

Meeting, 4 May 2017

- A & C think it would be fine to "publish" on Hardball Times, and would not preclude publishing for real, but are leary of putting my stuff out there before I write my package, and thus claim–and time stamp–my ideas as **my ideas.**
- Alix thinks I should write Nate Silver (who now works at ESPN, b/c they bought 538.com), introduce myself and my work a bit (further gussie up my website first), ask what work he could recommend.
- In the introduction of thesis there should be a lit *overview*, describing how the approaches to "Big N" came from other fields—because there was ZERO on this application, these were the major players and their articles (Gelman, Finley, Rue and Lindgren). Also, `ggplot2` and Shiny inspired and opened new doors... Hadley Wickham shout out.
  - RStan manual, Gelman book (and vignette);
  - Finley papers, `spBayes` manual;
  - SPDE papers, INLA papers, Lindstrom slides, `INLA` how-to paper;
  - Hadley's blogs, tutorials, package vignettes (ggplot2, dplyr);
  - MySQL book;
  - "Baseball Data with R" book, Cross-Sylvan;

  * Mention the almost complete scarcity of literature on baseball analysis with this rich new source of data.

  * Call the subsection "Research and Literature Overview"
- Okay to leave out labels/clutter on "choose your own resolution" portion.
- Alix doesn't want to get to a point where I am waiting on them to read my document.
- Alix **DOES want a passive voice sweep** before she reads it again.
- Former OSU student Louis Scott, who gave a talk to the math department a few weeks ago, started a website and is making boatloads of money.
- Alix has an idea for a new online/techy business, she's keeping it super secret (she told us, after making us swear to secrecy), and she hopes it's going to make her rich!
- Should probably find another example data set (ecology? Ethnicities? Crime dot maps?) for the Var-res HM chapter (and eventual package).
- If the plots are not in tip-top shape, then they won't look at plots when they read document.
- Got Lan's "Thesis Format" .tex docs, will slowly integrate them.
- Good idea to look at old dissertations in stats library, get a sense of things, how the introduction should look, for example.

1 8 May 2017

- Cool idea: a heat map of confidence interval widths should accompany original heat map of point estimates. :)
- Writing Hosmer-Lemeshow section. Took notes on first page of notebook. Downloaded appropriate chapter from their book.

2 Meeting

- Which photo of me for website? none of them, too informal.
- Heat map idea - eh.
- Color dissertation? Costs extra.
- mid-sentence citations? Rarely.
- USGS, Joe's role???
- "Faculty research assistant" (somebody's research grant)
- OSU job site??
- Samaritan???
- NuScale! (nuclear engineering)

May 18, 2017

- Confidence intervals for $\hat{p}$; in Myers et al. [2012], pg 143
- Recall, from Fall meeting notes:" Alix read my [first chapter], made some comments, said it was good. She thinks at the end there should be a section of sorts that helps the (baseball) reader understand and interpret the new map, rather than than just understanding the algorithm. What do the box sizes tell us? What does the particular spatial variation in box size tell us?"

May 22, 2017

"Personal R Packages"

https://hilaryparker.com/2013/04/03/personal-r-packages/

```
install_github('broman','kbroman')
library('broman')
```

"Note that `install_github` is a function in the `devtools` package."

"Writing an R package from scratch"

https://hilaryparker.com/2014/04/29/writing-an-r-package-from-scratch/

- (1) load `devtools` and `roxygen2`; (2) Create package directory; (3) add functions; (4) add documentation; (5) process documentation; (6) install; (7) Optional: add package as repository to GitHub

  1. Packages you will need, `devtools` and `roxygen2`.

     ```
     install.packages("devtools")
     library("devtools")
     devtools::install_github("klutometis/roxygen")
     library(roxygen2)
     ```

  2. Create package directory, bare minimum folders

     ```
     setwd("parent_directory")
     create("cats")
     ```

     This creates a folder called `cats` in the parent directory, with two folders (`man`, `R`) and a file (`DESCRIPTION`) inside.

  3. Add functions - add files containing the functions to the `R` folder (cats-package.r created automatically)

  4. Add documentation (with Superhero help: `roxygen2`). Example:

     ```
     #' A Cat Function
     #'
     #' This function allows you to express love of cats.
     #' @param love Do you love cats? Defaults to TRUE.
     #' @keywords cats
     #' @export
     #' @examples
     #' cat_function()

     cat_function <- function(love=TRUE){
         if(love==TRUE){print("I love cats!")}
         else {print("I am not a cool person.")}
     }
     ```

  5. Process your documentation (that you already created, previous step)

     ```
     setwd("./cats")
     document()
     ```

  6. Install

     ```
     setwd("..")
     install("cats")
     ```

  7. (Optional) Make the package a GitHub repo

     ```
     install_github('cats','github_username')
     ```

R Packages, by Hadley Wickham

- Package help: `package?x` and `help(package = "x")`
- "`devtools` - functions automate common development tasks

1. **Package Structure**
    - what `install.packages()` and `devtools::install_github()` do
    - Five package states: source, bundled, binary, installed, in-memory
    - ?? Load vs. attach vs. install ??
2. **R code**: most important directory is `R/`,
3. **Package metadata**: the `DESCRIPTION` lets you describe what your package needs to work.
4. **Documentation**: help people/you understand how to use the functions in your package, `roxygen2`
5. **Vignettes**: function documentation describes the nit-picky details of every function; Vignettes give the big picture.
6. **Tests**: to ensure your package works as designed
7. **Namespace**: functions it makes available, functions it requires from other packages.
8. **External data**: include data with your package.
9. **Compiled code**: ...compiled C and C++ code...
10. **Other components**: components that are rarely needed
11. **Git and github**
12. **Automated checking**
13. **Release**: release to the public, two main options (CRAN and github), general advice

Meeting

- For picture sizing, it seems like there are classes being used that are not defined in the obvious style.css files. This means it is probably pulling them from somewhere, like TwitterBootstrap or something. Need to find that to be able to choose.
- RMarkdown - link directly to the file, don't route through the GitHub viewer of the .pdf file. Use the `<a href="./lab7"` type of thing
- RMarkdown - you compile the .RMarkdown file into a .html file, instead of .pdf, then upload that to GitHub, and include the .rmd file in the repository
- Alix and Charlotte will come to party; will throw me a party after my defense
- Alix (and Charlotte) prefer not to read my paper until I would not make any more changes upon rereading it
- Alix and Charlotte prefer getting it in chunks
- A & C are in Corvallis over summer. Problem solved. Put a hard copy in their box, and let them know, when a chapter is ready.
- Data Analytics class — pittance
- Jasmine (African American —- lots of invites, etc)
- Not going to meet anymore, until further notice.

2.1 This week

- Git/github tutorial: http://kbroman.org/github_tutorial/
- Nice example personal webpage: http://alyssafrazee.com/
- Hot computer!! There was a *weird* process running (in triplicate!!), **fmesher**, which seems somehow related to Finn Lindgren, INLA, and github. WTF? The "CPU% usage" dropped from well above 90% to well under 10%; and it is cooling down!
  - "Try looking if it's a software conflict. In the **Other** folder in **Launchpad**, click **Activity Monitor**, this will list all the software processes running, look at the **%CPU** which has the biggest number running. Try to quit or stop that process. After a few moments, check if your MacBook is still hot. If it is not, then the process/software you stopped is the cause of the heat."
- Comprehensive R Archive Network, or CRAN
- Told Lee about my var-res heat maps, and Shiny CIs; she loved it! She said that with her data visuals tech experience, the var-res heat map would fundamentally change the way people present data. Fundamental change.
- The reaction I get from people about my var-res and ShinyCIs is so cool; I really should publish so people can see it, and use it.

Jasmine, Simple

- Jasmine's boss: "Director of Data Science"
- Seven "Data Scientists" at Simple;
- Three of the data scientists new (including Jasmine) as of late last year (October?)
- Quite a few more doing data science under other titles
- 330 employees total
- 1/3 of employees remote
- Four product "lines" (or groups, or divisions, or...)
- Three data scientist at Simple examples:
  - Jasmine - R, packages, computer science background; worked for Auto Insurance company
  - Neuroscience, research, communications
  - Computational math background
- Jasmine got the job: Customer → "Data science" position listed (was working for Auto Insurance company)
- Part of her role is mentoring; she repeatedly brought up "helping others" and ability to "coach others." Lots of working together, communicating, explaining things to one another.
- The "Onboarding" team is splitting up into "Risk" and "Onboarding"
- The assembled teams collectively cover the "Drew Conway - Venn Diagram" of statistician skills
- **Jasmine advice**: "don't get bogged down" trying to add specific skills; think of the Neuroscience data scientist example; you've put in a lot of time developing a depth of expertise...

- Jasmine says... I have an "awesome" background, and I definitely "have a shot" (to borrow my wording) at a Simple job
- Simple has a very learning, growing oriented culture; Jasmine spends two hours each morning on this part of herself/career/work
- Specific roles:
  - "Product Data Analyst" - "too Junior" for me
  - "Senior Data Analyst" - Industry specific knowledge, coach others, "risk modeling experience", new group/division so might need someone with training in this area to get the group started
  - New team: **"Decision Science"** (!!!!!!) - **Quantitative Analyst** - data science for business
- Applying to three roles is reasonable/acceptable
- Ask Jill Jubinski more about role specifics, which ones best for me, starting times, etc
  - "JillJubs" on Twitter
  - Also on LinkedIn
- Get on Twitter!! (#RStats) - Jasmine gave this as her "Biggest piece of advice" in response to a question at a recent talk
- Jasmine has a "remote friendly companies" forked repo on her GitHub page; said she would send it to me...
- Simple is based in Portland. So is Airbnb. Portland is (maybe?) a happy home for tech/remote companies.
- **Airbnb** has insane data science culture; 70 data scientists!
- **Stripe** - data engineer colleague just left for company called Stripe, to work remotely, Stripe based in Portland
- She said her tech skills weren't really appreciated or taken advantage of at AutoInsurance company; they didn't really know how or have capability
- Simple has a warehouse of cluster something or other, way ahead of banking curve for converting customer data to decisions/strategy.
- She said I can use her as a "reference" when I apply. :)
- She said it's not that hard to pick up new programming languages, but more important to be able to communicate, help others, explain ideas to other teams/members, mentor
- With Senior roles, "would never say don't apply"

13 June 2017

Package

- A peculiarity is blowing my mind. With three ball counts, there are a LOT of foul balls and swinging strikes—that do not affect BA.
- Number of swings by count:

```
> table(pitches$strike, pitches$ball)
        0      1      2      3
  0 170365 104989  39521   9915
  1 268712 291051 216331  92588
  2 205624 228970 167874  58544
```

- $\hat{p}$ by count (LOOK AT THREE BALL COUNTS!!!)

```
> rhbycount # right-handers by count
  Strikes  `0 Balls` `1 Ball` `2 Balls` `3 Balls`
        0    0.1977   0.1955    0.2048    0.0134
        1    0.1096   0.1016    0.0951    0.0194
        2    0.0205   0.0314    0.0426    0.0829
```

- Swing outcomes, ONLY three ball counts

```
> table(threeball$des)

                     Foul       Foul (Runner Going)                  Foul Tip
                   100727                      5310                      2217
          In play, no out           In play, out(s)           Swinging Strike
                     6785                     19248                     26113
Swinging Strike (Blocked)
                      647
```

- LOOK AT ALL THE FOUL BALLS!!! 100,000!!! 26,000 SWINGING STRIKES!!!
- What the fuck is going on? Are hitters overswinging!?!? ...but silently, b/c of so many foul balls and swinging strikes that do not affect their batting average???

3 June 21, 2017 - Summer Solstice!!

- Houston, we have a problem. For some reason, my package as I have it written now is eating a few observations each iteration. I don't know why. I am focusing on the first iteration, from one box to four, where I lost one observation. I think I need to record that puppy, and see which observation I am losing, see what its key attribute is, figure out how to fix it. I am using `as.image(...)`, which I think I have isolated as the step that loses the observation. Sigh.
- Oh yeah, I got a 0.6 FTE job offer from Reed. :)

4 Thursday, 29 June 2017

- Confidence intervals for $\hat{p}$ for Shiny app...
  1. $\widehat{\log\left(\frac{p}{1-p}\right)} \pm \text{m} * \text{SE}\left[\widehat{\log\left(\frac{p}{1-p}\right)}\right]$
  2. Back transform
  3. $\hat{p} \pm \text{M} * \text{SE}(\hat{p})$
- This is important b/c otherwise there can be CIs for $\hat{p}$ that include negative numbers, and thus incorrectly calibrated CI coverage.

- To check if $\widehat{\log\left(\frac{p}{1-p}\right)}$ is actually Normally distributed, a simulation study would proceed as follows.
  1. Fix $X\beta \rightarrow$ fix p.
  2. Generate Bernoulli(p) spatial data.
  3. Estimate everything; βs, thus logit(p), and thus p
  4. Iterate (2) and (3)
  5. Histograms: $\widehat{\text{logit}(p)}$ vs. $\hat{p}$

As of Now

- Dissertation - chip away
- Make packages
- Webpage
  - Talks (JSM, seminar)
  - RMarkdown, ST516 & ST517 handouts
  - Consulting reports/summaries
- Graduation and commencement

References

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(4):423–498, 2011.

Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. Journal of Statistical Software, 63 (19), 2015.

Daniel Simpson, Finn Lindgren, and Håvard Rue. Think continuous: Markovian gaussian models in spatial statistics. Spatial Statistics, 1:16–29, 2012a.

Daniel Simpson, Finn Lindgren, and Håvard Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. Environmetrics, 23(1):65–74, 2012b.

Oliver Schabenberger and Carol A Gotway. Statistical methods for spatial data analysis. CRC press, 2004.

Xuerong Mao. Stochastic differential equations and applications. Elsevier, 2007.

Johan Lindstrom, Finn Lindgren, and Havard Rue. Gaussian markov random fields: Efficient modelling of spatially dependent data. Lecture Slides, 2011. Centre for Mathematical Sciences, Lund University.

Johan Lindstrom. Gaussian markov random fields. Lecture Slides, 2014. Pan-American Advanced Study Institute, Buzios.

Johan Lindstrom. Latent gaussian processes and stochastic partial differential equations. Lecture Slides, 2016. Centre for Mathematical Sciences, Lund University.

HÅvard Rue and Sara Martino. Approximate bayesian inference for hierarchical gaussian markov random field models. Journal of statistical planning and inference, 137(10):3177–3192, 2007.

Havard Rue and Leonhard Held. Gaussian Markov random fields: theory and applications. CRC Press, 2005.

Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. Journal of the american statistical association, 81(393):82–86, 1986.

Raymond H Myers, Douglas C Montgomery, G Geoffrey Vining, and Timothy J Robinson. Generalized linear models: with applications in engineering and the sciences, volume 791. John Wiley & Sons, 2012.