Spatial GLMs

Generalized Linear Model Notation

There are three components to a generalized linear model:

1. Sampling Distribution: such as Poisson or Binomial

2. Linear combination of predictors: $\eta = X\beta$

3. A link function to map the linear combination of predictors to the support of the sampling distribution.

Binary Regression Overview

Write out the complete model specification for binary regression.

- Assume Y_i is the binary response for the i^{th} observation,

$$\begin{aligned} Y_i &\sim Bernoulli(\pi_i) \\ logit(\pi_i) &= X_i\beta, \\ \text{or } \Phi^{-1}(\pi_i) &= X_i\beta \end{aligned}$$

- where $logit(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right)$

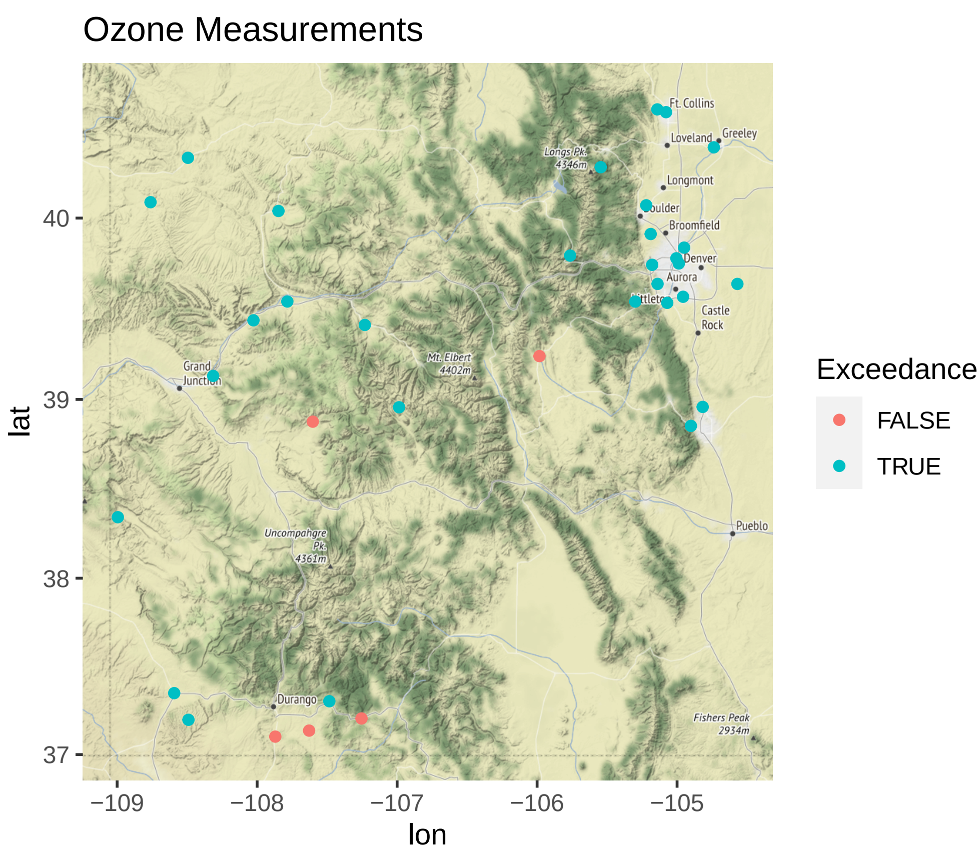
- where $\Phi() =$ is the CDF of a standard normal distribution

Latent interpretation of probit model:

Let $z_i > 0$ if $y_i = 1$. Otherwise, let $z_i < 0$. Then $z_i \sim N(X\beta, 1)$ is a latent continuous variable that is mapped to zero or 1.

For a set of predictors, $X^{'}$, then $z^{'} \sim N(X^{'}\beta, 1)$ and the probability of a 1 or zero can be obtained by integrating the latent distribution.

Consider air quality data from Colorado as a motivating example.



Interpret the output.

```
CO <- CO %>% mutate(north = as.numeric(Latitude > 38 ))
glm(Exceedance~north, family=binomial(link = 'probit'),data=CO) %>% display()
```

```
## glm(formula = Exceedance ~ north, family = binomial(link = "probit"),
##     data = CO)
##             coef.est coef.se
## (Intercept) 0.00     0.51
## north       1.48     0.62
## ---
##   n = 35, k = 2
##   residual deviance = 22.9, null deviance = 28.7 (difference = 5.8)
```

```
glm(Exceedance~north, family=binomial(link = 'logit'),data=CO) %>% display()
```

```
## glm(formula = Exceedance ~ north, family = binomial(link = "logit"),
##     data = CO)
##             coef.est coef.se
## (Intercept) 0.00     0.82
## north       2.60     1.10
## ---
##   n = 35, k = 2
##   residual deviance = 22.9, null deviance = 28.7 (difference = 5.8)
```

Spatial Binary Regression

Assume $Y(\boldsymbol{s_i})$ is the binary response for $\boldsymbol{s_i}$,

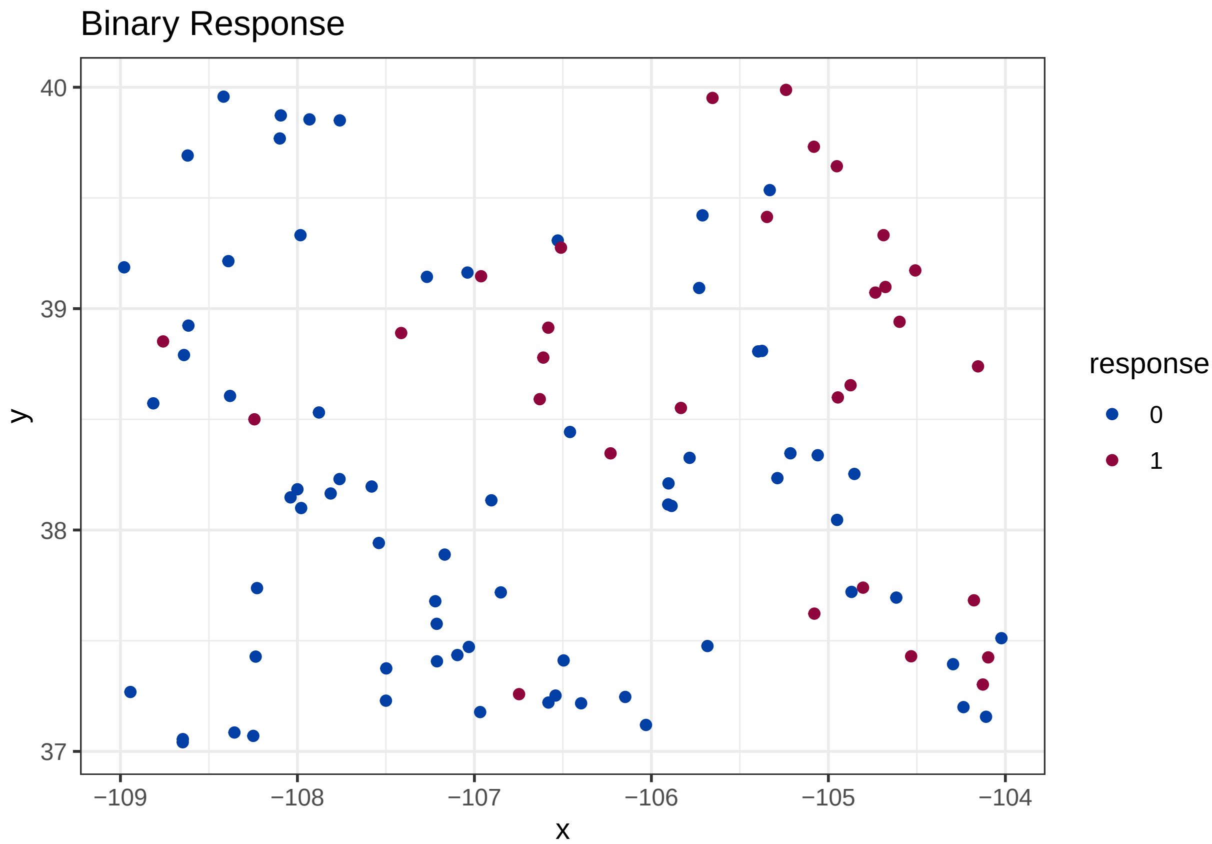
$$\begin{aligned} Y(\boldsymbol{s_i})|\beta, w(\boldsymbol{s_i}) &\sim Bernoulli(\pi(\boldsymbol{s_i})) \\ \Phi^{-1}(\pi(\boldsymbol{s_i})) &= X(\boldsymbol{s_i})\beta + w(\boldsymbol{s_i}), \end{aligned}$$

where $\boldsymbol{W} \sim N(\boldsymbol{0}, \sigma^2 H(\phi))$

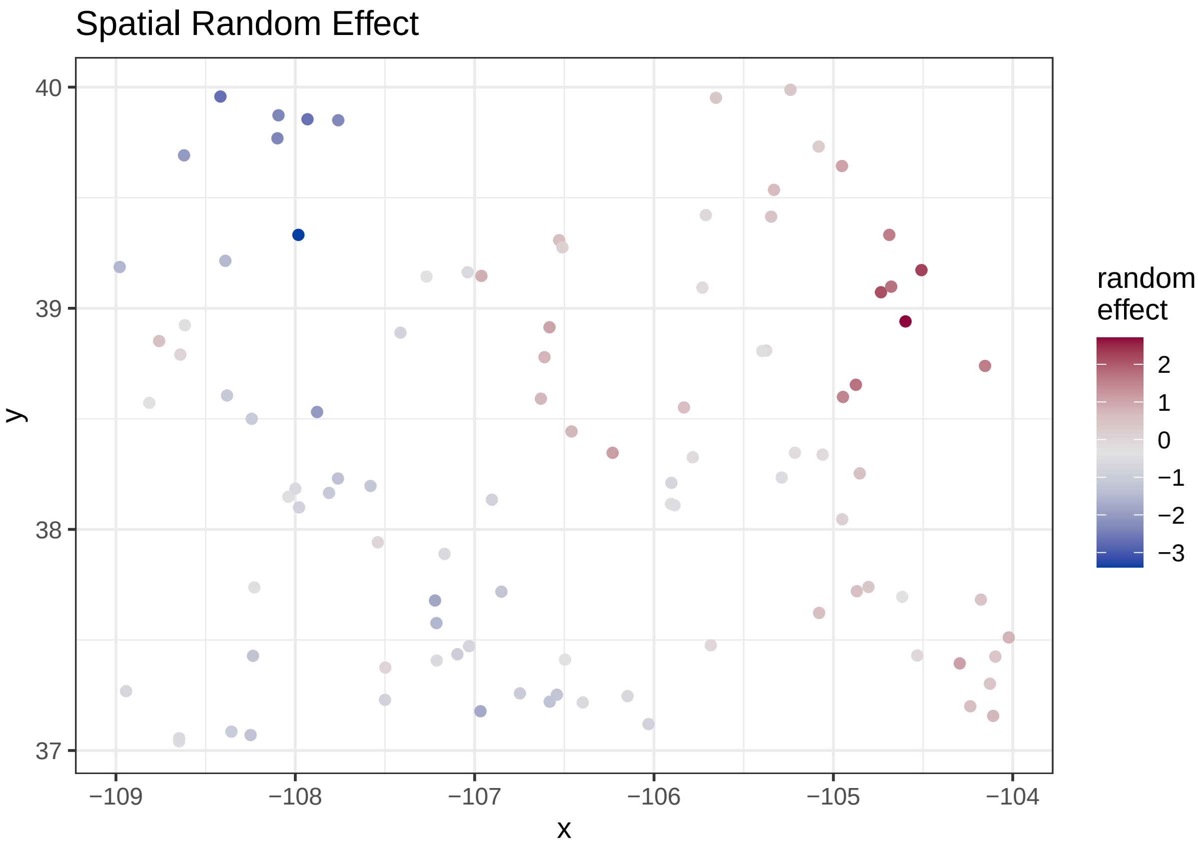
Simulating spatial random effects for binary data

```r
N.sim <- 100
Lat.sim <- runif(N.sim,37,40)
Long.sim <- runif(N.sim,-109,-104)
phi.sim <- 1
sigmasq.sim <- 1
beta.sim <- c(-1,1)
north.sim <-  as.numeric(Lat.sim > 38)


d <- dist(cbind(Lat.sim,Long.sim), upper = T, diag = T) %>% as.matrix
H.sim <- sigmasq.sim * exp(- d / phi.sim)
w.sim <- rmnorm(1,0,H.sim)
xb.sim <- beta.sim[1] + beta.sim[2] * north.sim
y.sim <- rbinom(N.sim,1,pnorm(xb.sim + w.sim))
```



Binary Response
40
39
38
37
y
−109
−108
−107
−106
−105
−104
x
response
0
1



Spatial Random Effect
40
39
38
37
y
−109
−108
−107
−106
−105
−104
x
random effect
2
1
0
−1
−2
−3

STAN: probit regression

```
##
## // The input data is a vector 'y' of length 'N'.
## data {
##   int<lower=0> N;
##   int<lower=0,upper=1> y[N];
##   vector[N] x;
## }
##
## // The parameters accepted by the model. Our model
## // accepts two parameters 'mu' and 'sigma'.
## parameters {
##   real beta0;
##   real beta1;
## }
##
## transformed parameters {
##   real<lower = 0, upper = 1> p[N];
##   for (i in 1:N) {
##     p[i] = Phi(beta0 + beta1 * x[i]);
##   }
## }
##
##
## // The model to be estimated.
## model {
##   for (i in 1:N){
##     y[i] ~ bernoulli(p[i]);
##   }
## }
```

Binary Regression

```
probit_stan <- stan(file = 'probit_regression.stan', data = list(N = N.sim, y = y.sim, x = north.sim))
```

```
print(probit_stan, pars = c('beta0', 'beta1'))
```

```
## Inference for Stan model: probit_regression.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##        mean se_mean   sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
## beta0 -0.92    0.01 0.24 -1.40 -1.08 -0.91 -0.76 -0.48   975    1
## beta1  0.59    0.01 0.29  0.04  0.39  0.59  0.79  1.18   985    1
##
## Samples were drawn using NUTS(diag_e) at Fri Mar 12 08:55:30 2021.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```
glm(y.sim ~ north.sim, family = binomial(link = 'probit'))
```

```
##
## Call:  glm(formula = y.sim ~ north.sim, family = binomial(link = "probit"))
```

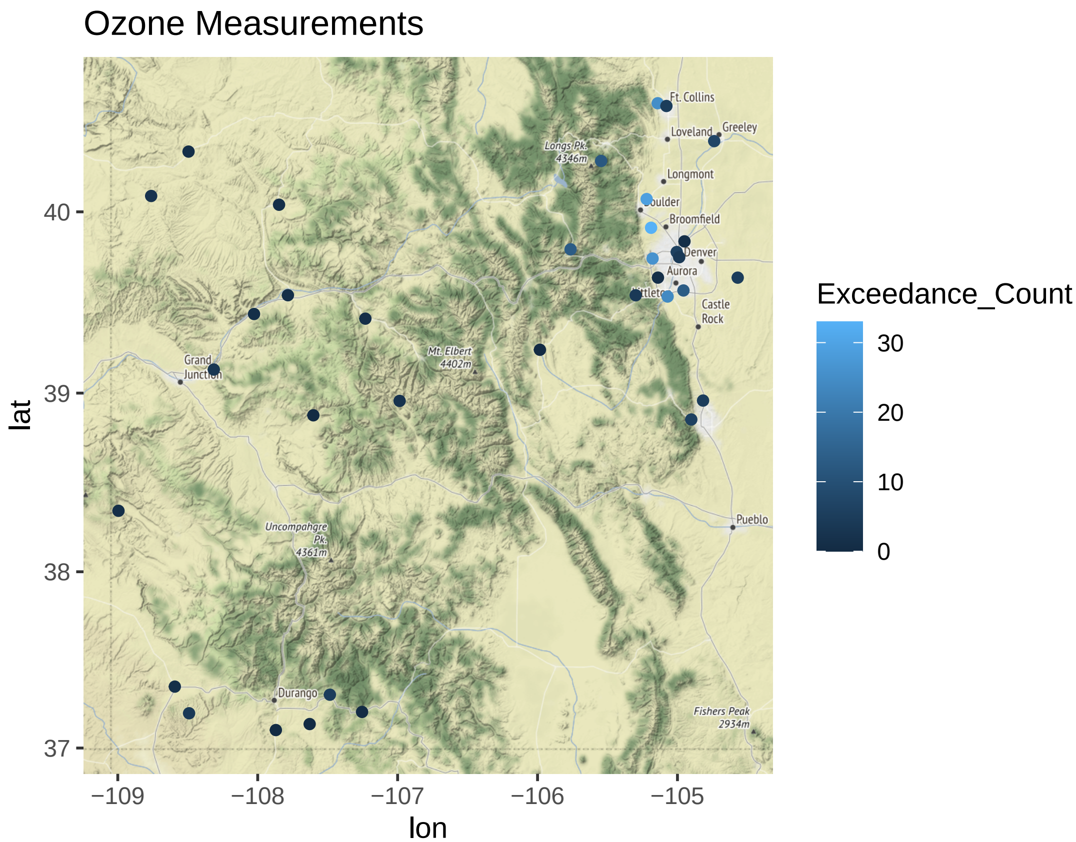
```
##
## Coefficients:
## (Intercept)    north.sim
##     -0.8994       0.5701
##
## Degrees of Freedom: 99 Total (i.e. Null);  98 Residual
## Null Deviance:       122.2
## Residual Deviance: 118.1     AIC: 122.1
```

```
tibble(y.sim = y.sim, north.sim = north.sim) %>% stan_glm(y.sim ~ north.sim, family = binomial(link = 'p
```

```
## stan_glm
##  family:       binomial [probit]
##  formula:      y.sim ~ north.sim
##  observations: 100
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept) -0.9    0.2
## north.sim    0.6    0.3
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Spatial Poisson Regression

Motivation



Poisson Regression Overview

Write out the complete model specification for Poisson regression.

Assume Y_i is the count response for the i^{th} observation,

$$\begin{aligned} Y_i &\sim Poisson(\lambda_i) \\ \log(\lambda_i) &= X_i\beta, \end{aligned}$$

thus $\exp(X_i\beta) \geq 0$

Next write out a Poisson regression model with spatial random effects

$$\begin{aligned} Y(\boldsymbol{s_i}) &\sim Poisson(\lambda(\boldsymbol{s_i})) \\ \log(\lambda(\boldsymbol{s_i})) &= X(\boldsymbol{s_i})\beta + w(\boldsymbol{s_i}), \end{aligned}$$

where $\boldsymbol{W} \sim N(\boldsymbol{0}, \sigma^2 H(\phi))$