Lecture 1

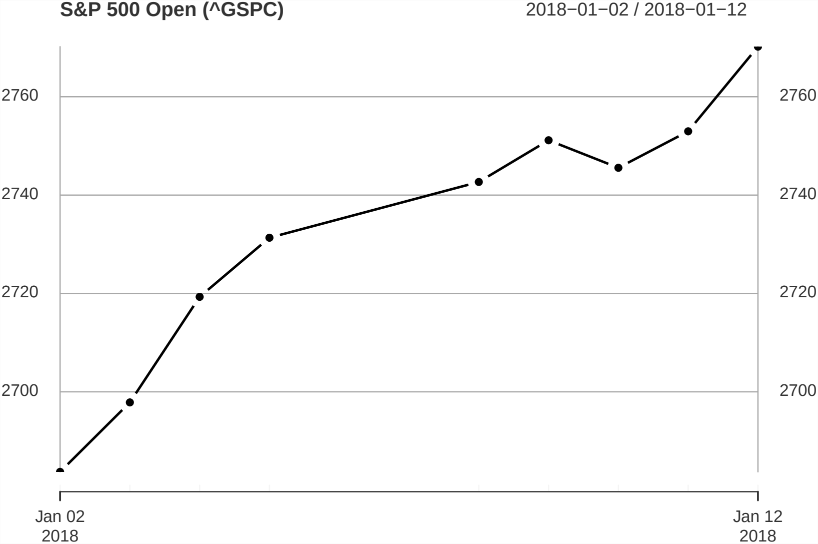
Spatio-temporal data & Linear Models

Colin Rundel

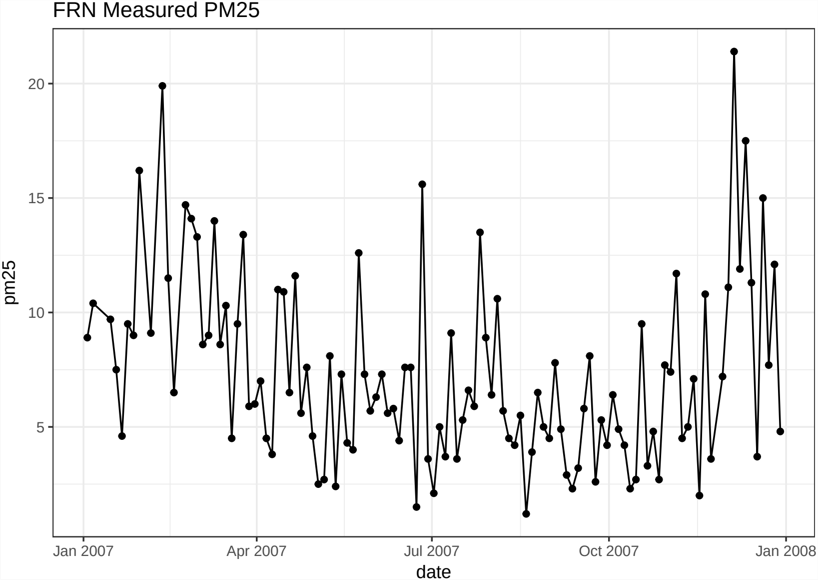
1/16/2018

Spatio-temporal data

Time Series Data - Discrete

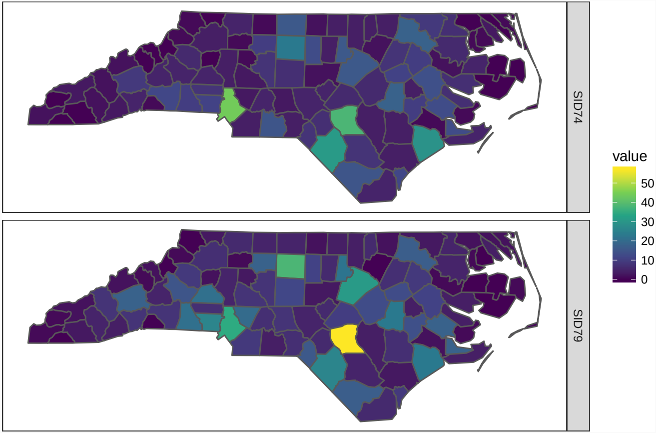


Time Series Data - Continuous

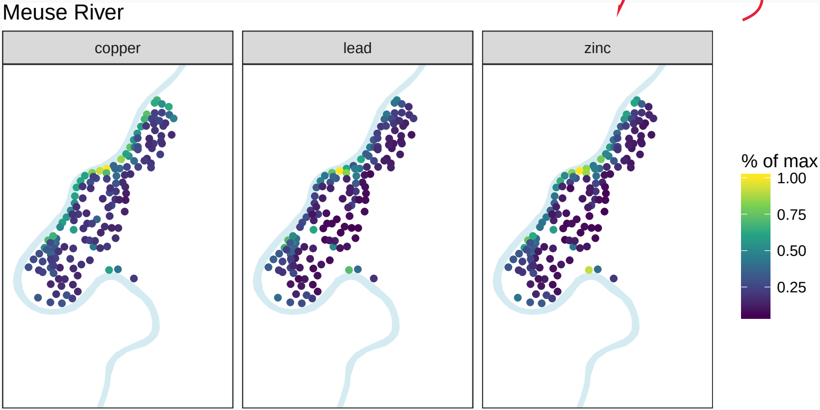


Spatial Data - Areal

infant death

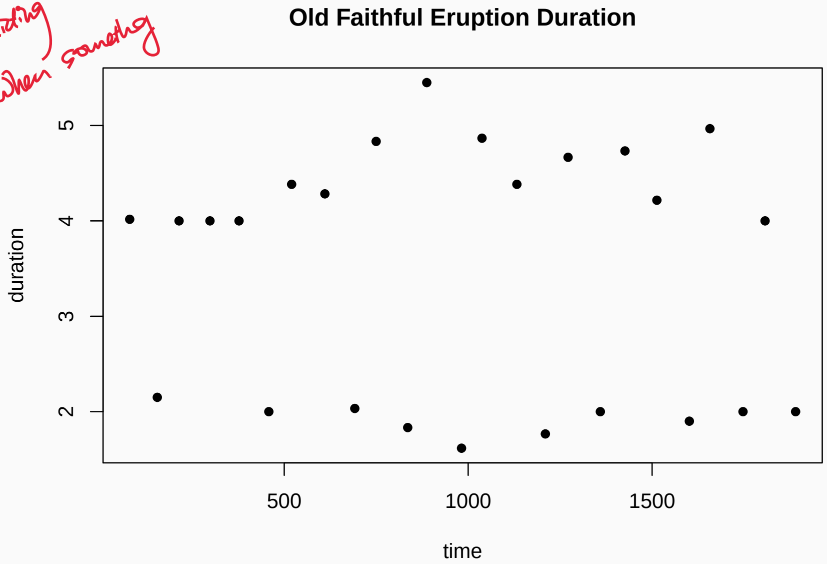


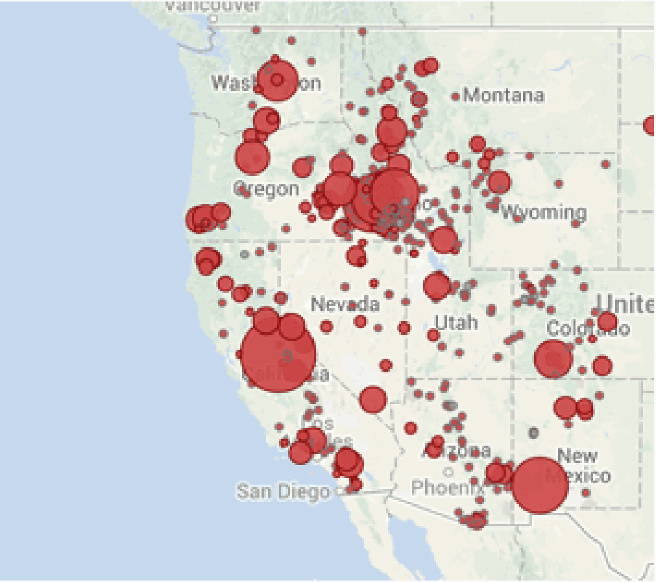
chemical company dumping



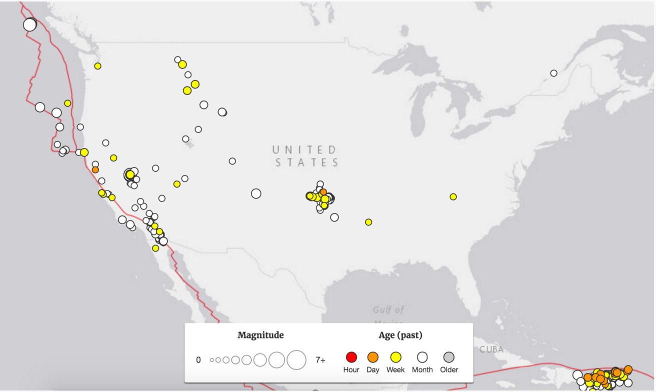
Point Pattern Data - Time

predicting when something




Vancouver
Montana
Oregon
Wyoming
Nevada
Utah
Colorado
New Mexico
San Diego
Phoenix

when. where. how intense



Simple, IID
$\Rightarrow$
IID fail.

(Bayesian) Linear Models

Pretty much everything we a going to see in this course will fall under the umbrella of linear or generalized linear models.

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1\, x_{i1} + \cdots + \beta_p\, x_{ip} + \epsilon_i \\
\epsilon_i &\sim N(0, \sigma^2)
\end{aligned}
$$

which we can also express using matrix notation as

$$
\begin{aligned}
\underset{n\times 1}{\mathbf{Y}} &= \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}} \\
\boldsymbol{\epsilon} &\sim N(\underset{n\times 1}{\mathbf{0}},\ \sigma^2\,\underset{n\times n}{\mathbb{1}_n})
\end{aligned}
$$

→ 把Dimension写出来有好处

Multivariate Normal Distribution

For an n-dimension multivate normal distribution with covariance $\boldsymbol{\Sigma}$ (positive semidefinite) can be written as

$$\underset{n\times 1}{\mathbf{Y}} \sim N(\underset{n\times 1}{\boldsymbol{\mu}}, \underset{n\times n}{\boldsymbol{\Sigma}}) \text{ where } \{\boldsymbol{\Sigma}\}_{ij} = \rho_{ij}\sigma_i\sigma_j$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \rho_{11}\sigma_1\sigma_1 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \rho_{nn}\sigma_n\sigma_n \end{pmatrix}\right)$$

Computation

Never directly invert a matrix!

Big data

$>$ 10,000 obs $\Rightarrow$ get slow

n^3.

Multivariate Normal Distribution - Density

For the n dimensional multivate normal given on the last slide, its density is given by

$$(2\pi)^{-n/2} \ \det(\boldsymbol{\Sigma})^{-1/2} \ \exp\left(-\frac{1}{2}(\underset{1\times n}{\mathbf{Y}-\boldsymbol{\mu}})'\underset{n\times n}{\boldsymbol{\Sigma}^{-1}}(\underset{n\times 1}{\mathbf{Y}-\boldsymbol{\mu}})\right)$$

?

and its log density is given by

$$-\frac{n}{2}\log 2\pi - \frac{1}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\underset{1\times n}{\mathbf{Y}-\boldsymbol{\mu}})'\underset{n\times n}{\boldsymbol{\Sigma}^{-1}}(\underset{n\times 1}{\mathbf{Y}-\boldsymbol{\mu}})$$

"The Matrix cookbook" ~~reference book~~. will provide link.

$$\log(\beta, \sigma^2 | X, y) = -\frac{n}{2}\cancel{\log 2\pi} - \frac{1}{2}\det\Sigma - \frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)$$

$$\Sigma = \sigma^2 \mathbb{1}_n = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{pmatrix}$$

$$\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial l(\beta, \sigma^2 | x, y)}{\partial\beta} \propto \left[(-X)'(y - X\beta) + (-X)'(y - X\beta)\right]$$

$$\propto -2\underset{p\times n}{X'}(\underset{n\times 1}{y} - \underset{n\times p}{X}\underset{p\times 1}{\beta}) \stackrel{\text{let}}{=} 0$$

$\boxed{\text{dim}}$

$$X'(y - X\beta) = 0 \Rightarrow X'y - X'X\beta = 0$$

$$\boxed{(X'X)^{-1}X'y = \hat{\beta}_{MLE}}$$

Maximum Likelihood - σ^2

$$\frac{\partial \ell(\beta, \sigma^2 | X, y)}{\partial \sigma^2} \propto \frac{-n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)'(y - X\beta)$$

$$\propto \frac{1}{2\sigma^2}\left(-n + \frac{1}{\sigma^2}(y - X\beta)'(y - X\beta)\right)$$

Assume $\sigma^2 > 0$

$$\sigma^2 = \frac{1}{n}(y - X\beta)'(y - X\beta)$$

$$= \frac{1}{n}\sum(y_i - X_i\beta)^2 = \frac{1}{n}\sum(y - \hat{y}_i)^2$$

Likelihood:

$$\mathbf{Y} \mid \boldsymbol{\beta},\, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta},\, \sigma^2\, \mathbb{1}_n)$$

Likelihood:

$$\mathbf{Y} \mid \boldsymbol{\beta},\, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta},\, \sigma^2\, \mathbb{1}_n)$$

Priors:

$$\beta_i \sim N(0, \sigma_\beta^2) \text{ or } \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2\, \mathbb{1}_{\boldsymbol{p}})$$

$$\sigma^2 \sim \text{Inv-Gamma}(a,\, b)$$

Deriving the posterior

$$\begin{aligned}
[\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{Y}, \mathbf{X}] &= \frac{[\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2]}{[\mathbf{Y} \,|\, \mathbf{X}]} [\boldsymbol{\beta}, \sigma^2] \\
&\propto [\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta}, \, \sigma^2] \\
&\propto [\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta} \,|\, \sigma^2][\sigma^2]
\end{aligned}$$

Deriving the posterior

$$
\begin{aligned}
[\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{Y}, \mathbf{X}] &= \frac{[\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2]}{[\mathbf{Y} \,|\, \mathbf{X}]}[\boldsymbol{\beta}, \sigma^2] \\
&\propto [\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta},\, \sigma^2] \\
&\propto [\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta} \,|\, \sigma^2][\sigma^2]
\end{aligned}
$$

where,

$$
f(\mathbf{Y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)
$$

Deriving the posterior

$$
\begin{aligned}
[\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}, \mathbf{X}] &= \frac{[\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2]}{[\mathbf{Y} \mid \mathbf{X}]}[\boldsymbol{\beta}, \sigma^2] \\
&\propto [\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta},\ \sigma^2] \\
&\propto [\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta} \mid \sigma^2][\sigma^2]
\end{aligned}
$$

where,

$$
f(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)
$$

$$
f(\boldsymbol{\beta} \mid \sigma^2_\beta) = (2\pi\sigma^2_\beta)^{-p/2} \exp\left(-\frac{1}{2\sigma^2_\beta}\boldsymbol{\beta}'\boldsymbol{\beta}\right)
$$

Deriving the posterior

$$
\begin{aligned}
[\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}, \mathbf{X}] &= \frac{[\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2]}{[\mathbf{Y} \mid \mathbf{X}]}[\boldsymbol{\beta}, \sigma^2] \\
&\propto [\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta},\ \sigma^2] \\
&\propto [\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta} \mid \sigma^2][\sigma^2]
\end{aligned}
$$

where,

$$
f(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)
$$

$$
f(\boldsymbol{\beta} \mid \sigma^2_\beta) = (2\pi\sigma^2_\beta)^{-p/2} \exp\left(-\frac{1}{2\sigma^2_\beta}\boldsymbol{\beta}'\boldsymbol{\beta}\right)
$$

$$
f(\sigma^2 \mid a,\ b) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)
$$

Deriving the Gibbs sampler (σ^2 step)

$$[\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}, \mathbf{X}] \propto (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)$$

$$(2\pi\sigma_\beta^2)^{-p/2} \exp\left(-\frac{1}{2\sigma_\beta^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right)$$

$$\frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)$$

$s = \frac{\sigma^2}{\sigma_\beta^2}$

$\sigma^2 | \cdots \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta})'(Y - \underset{\sim}{X}\underset{\sim}{\beta})\right.$

$(\sigma^2)^{-1} \exp\left(-\frac{1}{\sigma^2} b\right)$

$\propto (\sigma^2)^{-(a+n/2)-1} \exp\left(-\frac{1}{\sigma^2} \cdots\right)$

$\therefore \quad \sigma^2 | \cdots \sim$ Inv-Gamma

$$[\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}, \mathbf{X}] \propto \cancel{(2\pi\sigma^2)}^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)$$

$$\cancel{(2\pi\sigma_\beta^2)^{-p/2}} \exp\left(-\frac{1}{2\sigma_\beta^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right) \qquad S = \frac{\sigma^2}{\sigma_\beta^2}$$

$$\cancel{\frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)} \qquad \exp\left(-\frac{S}{2\sigma^2}\beta'\beta\right)$$

$$\beta \mid \cdot \propto \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta}) - \frac{S}{2\sigma^2}\beta'\beta\right\}.$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left(y'y - \beta'X'y - y'X\beta - \beta'X'X\beta - \beta'S\beta\right)\right\}$$

Know $\beta \sim N(\mu_p, \Sigma_p)$

$$\beta \mid \cdot \propto \exp\left[-\frac{1}{2}(\beta-\mu_p)'\Sigma_p^{-1}(\beta-\mu_p)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\beta'\Sigma_p^{-1}\beta - \beta'\Sigma_p^{-1}\mu_p \cdots\right)\right]$$

$$\Sigma_p^{-1} = (X'X + S\mathbb{I}) \qquad \Sigma_p = (X'X + S\mathbb{I})^{-1}$$

$$= \left(X'X + \frac{\sigma^2}{\sigma_\beta^2}\mathbb{I}\right)^{-1}$$

$$\mu_p = -\beta' X' y \Rightarrow \mu_p = \Sigma_p (X'y).$$

$$= \left(X'X + \frac{\sigma^2}{\sigma_p^2}\mathbb{I}\right)^{-1} X'y.$$

从结果套型.

把 μ_p, Σ_p 对应前面的(上页)乱七八糟结果. 可解.

A Quick Example

Some Fake Data

Lets generate some simulated data where the underlying model is known and see how various regression preceedures function.

$$\beta_0 = 0.7, \quad \beta_1 = 1.5, \quad \beta_2 = -2.2, \quad \beta_3 = 0.1$$
$$n = 100, \quad \epsilon_i \sim N(0,1)$$

Generating the data

```r
set.seed(01162018)
n = 100
beta = c(0.7,1.5,-2.2,0.1)
eps = rnorm(n)

d = data_frame(
  X1 = rt(n,df=5),
  X2 = rt(n,df=5),
  X3 = rt(n,df=5)
) %>%
  mutate(Y = beta[1] + beta[2]*X1 + beta[3]*X2 + beta[4]*X3 + eps)

X = cbind(1, d$X1, d$X2, d$X3)
```

Least squares fit

Let $\hat{\mathbf{Y}}$ be our estimate for $\mathbf{Y}$ based on our estimate of $\boldsymbol{\beta}$,

$$\hat{\mathbf{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \, \mathbf{X}_1 + \hat{\beta}_2 \, \mathbf{X}_2 + \hat{\beta}_3 \, \mathbf{X}_3 = \mathbf{X} \, \hat{\boldsymbol{\beta}}$$

Let $\hat{\mathbf{Y}}$ be our estimate for $\mathbf{Y}$ based on our estimate of $\boldsymbol{\beta}$,

$$\hat{\mathbf{Y}} = \hat{\beta}_0 + \hat{\beta}_1\,\mathbf{X}_1 + \hat{\beta}_2\,\mathbf{X}_2 + \hat{\beta}_3\,\mathbf{X}_3 = \mathbf{X}\,\hat{\boldsymbol{\beta}}$$

The least squares estimate, $\hat{\boldsymbol{\beta}}_{ls}$, is given by

$$\underset{\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \mathbf{X}_{i\cdot}\boldsymbol{\beta}\right)^2$$

Let $\hat{\mathbf{Y}}$ be our estimate for $\mathbf{Y}$ based on our estimate of $\boldsymbol{\beta}$,

$$\hat{\mathbf{Y}} = \hat{\beta}_0 + \hat{\beta}_1\,\mathbf{X}_1 + \hat{\beta}_2\,\mathbf{X}_2 + \hat{\beta}_3\,\mathbf{X}_3 = \mathbf{X}\,\hat{\boldsymbol{\beta}}$$

The least squares estimate, $\hat{\boldsymbol{\beta}}_{ls}$, is given by

$$\underset{\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \mathbf{X}_{i\cdot}\boldsymbol{\beta}\right)^2$$

Previously we derived,

$$\hat{\boldsymbol{\beta}}_{ls} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{Y}$$

Frequentist Fit

```
l = lm(Y ~ X1 + X2 + X3, data=d)
l$coefficients
```

```
## (Intercept)          X1          X2          X3
##   0.6566561   1.4657537  -2.2807109   0.1629704
```

```
(beta_hat = solve(t(X) %*% X, t(X)) %*% d$Y)
```

```
##            [,1]
## [1,]  0.6566561
## [2,]  1.4657537
## [3,] -2.2807109
## [4,]  0.1629704
```

Bayesian model specification (JAGS)

★ JAGS Note:
tau is precision, not variance

```
model =
"model{
  # Likelihood
  for(i in 1:length(Y)){
    Y[i]   ~ dnorm(mu[i], tau)
    mu[i] = beta[1] + beta[2]*X1[i] + beta[3]*X2[i] + beta[4]*X3[i]
  }

  # Prior for beta
  for(j in 1:4){
    beta[j] ~ dnorm(0,1/100)
  }

  # Prior for sigma / tau2
  tau ~ dgamma(1, 1)
  sigma2 = 1/tau
}"
```

Bayesian model fitting (JAGS)

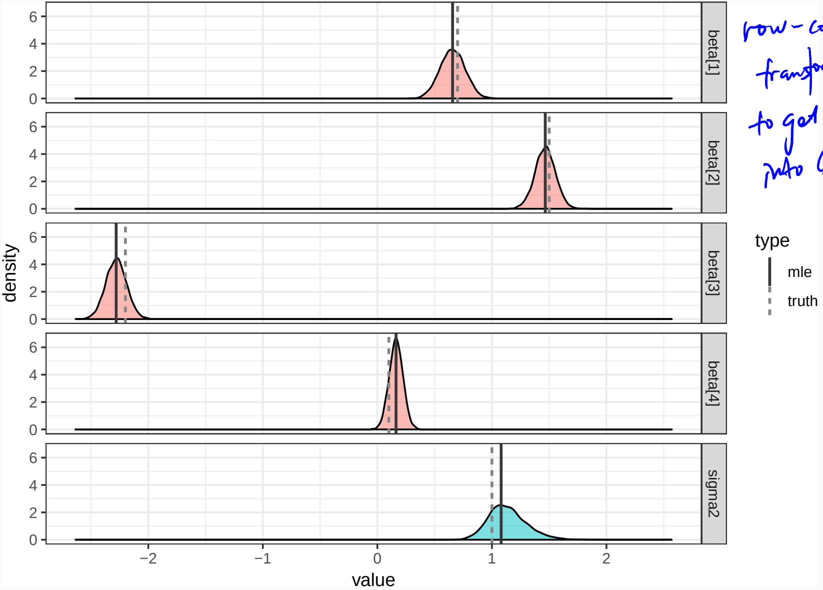
```
m = rjags::jags.model(
  textConnection(model),
  data = d
)
```

a model file → pretend "model" is a file

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 100
##    Unobserved stochastic nodes: 5
##    Total graph size: 810
##
## Initializing model
```

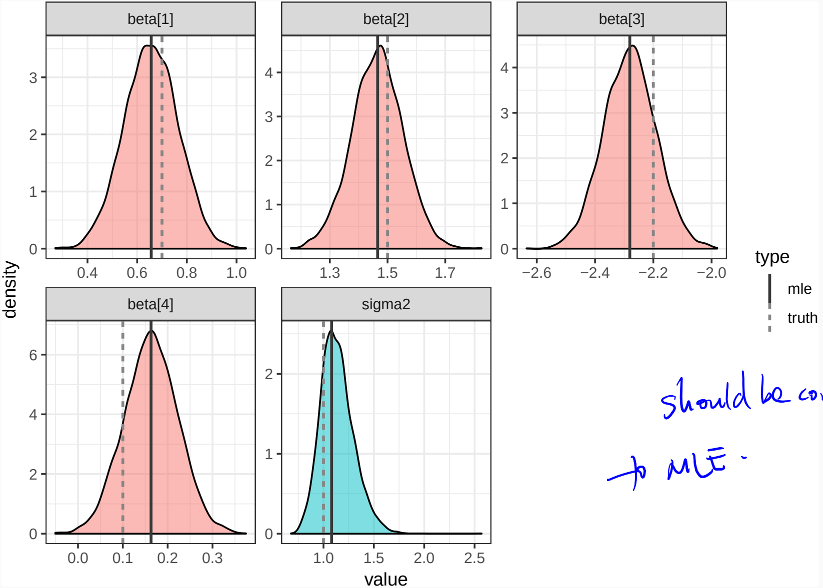
```
update(m, n.iter=1000, progress.bar="none")

samp = rjags::coda.samples(
  m, variable.names=c("beta","sigma2"),
  n.iter=5000, progress.bar="none"
)
```

Results



row-column transformation to get it into GGplot

Results (zoom)



should be converging to MLE.