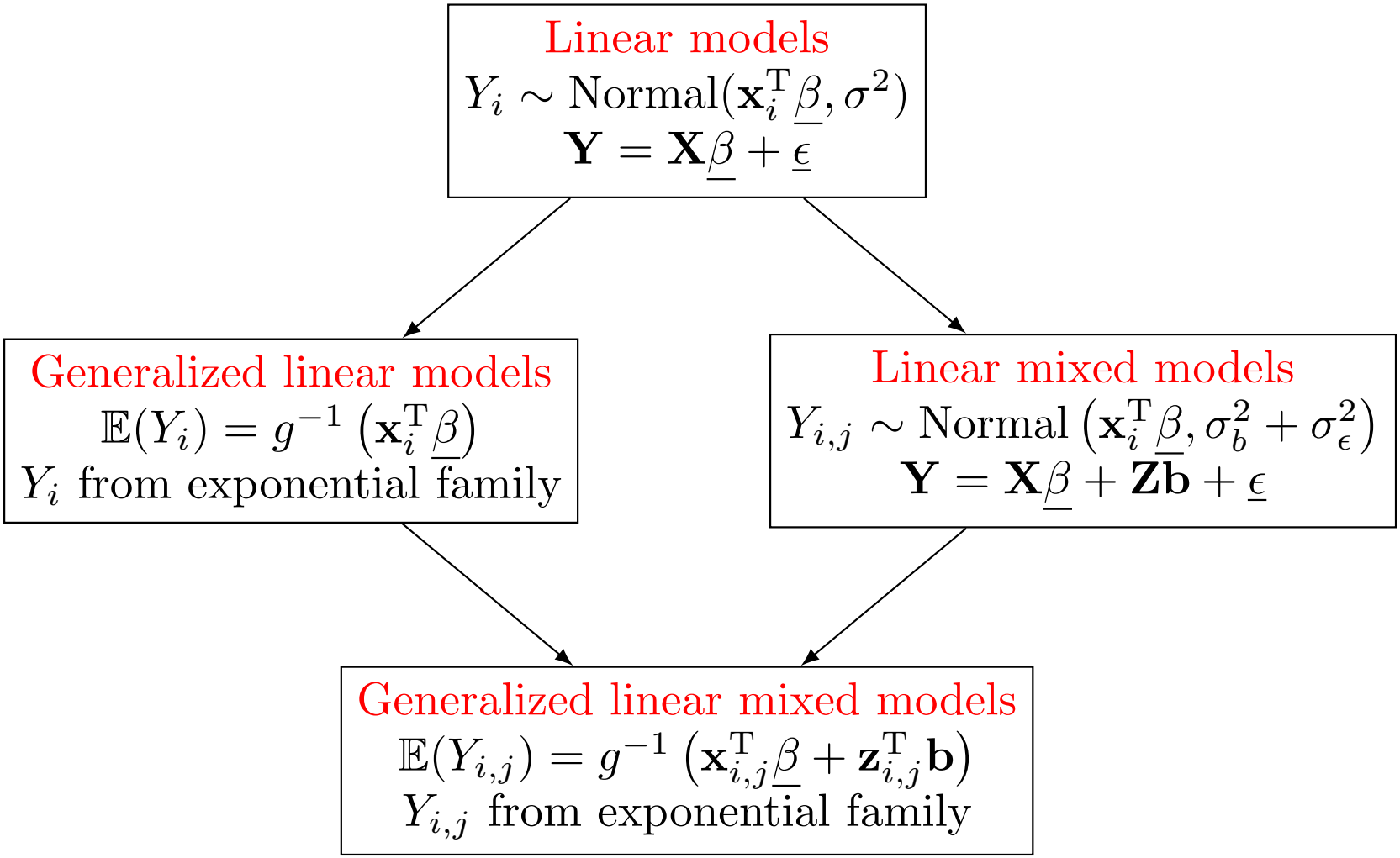
MA50260 Statistical Modelling

Lecture 19: GLMMs - Inference and Model Selection

Ilaria Bussoli

April 23, 2024

Generalised Linear Mixed Models



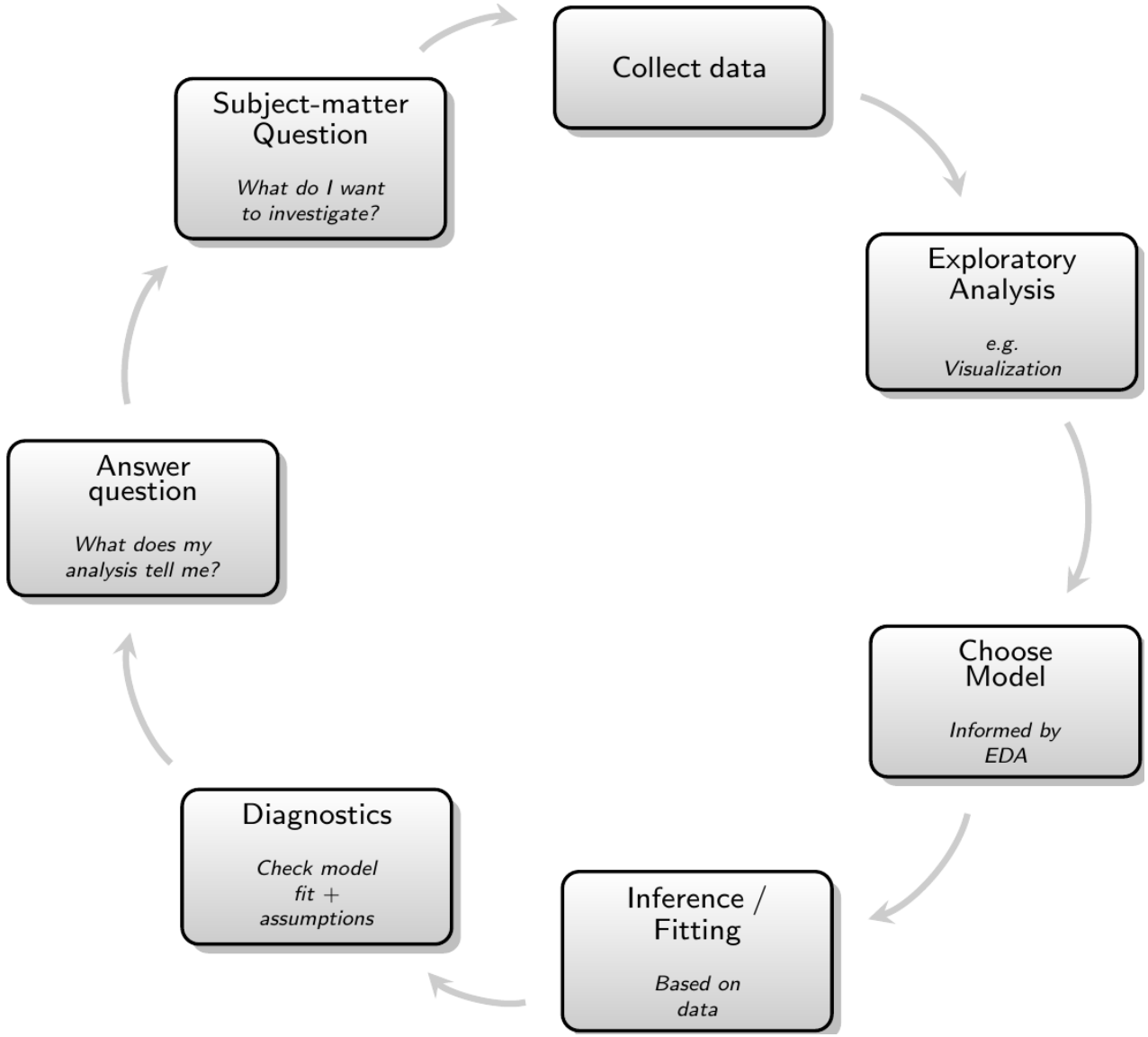
GLMMs - Estimation

In the last lecture, we considered two approaches for estimating GLMMs:

- Penalized quasi likelihood
- Gauss-Hermite quadrature

How do these methods compare?

Philosophy of Statistical Modelling



Sampling distribution of $\hat{\underline{\beta}}$

For a GLM, we have

$$\hat{\underline{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_{\boldsymbol{p}}\left(\underline{\beta}\,,\, \phi(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\right).$$

Further, for a (normal) linear mixed model,

$$\hat{\underline{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_{\boldsymbol{p}}(\underline{\beta}, (\mathbf{X}^{\mathrm{T}}(\mathbf{Z}\mathcal{D}\mathbf{Z}^{\mathrm{T}} + \sigma_{\epsilon}^{2}\mathbf{I}_{n})^{-1}\mathbf{X})^{-1}).$$

Sampling distribution of $\hat{\underline{\beta}}$

For a GLM, we have

$$\hat{\underline{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_{\boldsymbol{p}}\left(\underline{\beta}\,,\, \phi(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\right).$$

Further, for a (normal) linear mixed model,

$$\hat{\underline{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_{\boldsymbol{p}}(\underline{\beta}, (\mathbf{X}^{\mathrm{T}}(\mathbf{Z}\mathcal{D}\mathbf{Z}^{\mathrm{T}} + \sigma_{\epsilon}^{2}\mathbf{I}_{\boldsymbol{n}})^{-1}\mathbf{X})^{-1}).$$

The sampling distribution for the GLMM is

$$\hat{\underline{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_{\boldsymbol{p}}(\underline{\beta}, (\mathbf{X}^{\mathrm{T}}(\mathbf{Z}\mathcal{D}\mathbf{Z}^{\mathrm{T}} + \mathbf{W}^{-1}\phi)^{-1}\mathbf{X})^{-1}).$$

Can you show that this result agrees with the results above?

Example - Seed Germination

We estimate the model defined in Lecture 17

```
load("seeds.rda")
estim <- glmer(
    cbind(Germinated,NotGerminated) ~ (1 | Plate),
    nAGQ = 25, data = seeds, family = binomial
)
```

Let's consider the estimated fixed effect

```
round( summary(estim)$coefficients, 3 )
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.377      0.263   -1.43    0.153
```

How can we give a 95% confidence interval for β_1?

Exercise - Ohio Wheeze Data

Let's consider the results for the Ohio wheeze data with 536 children

```
round( summary( estimGH )$coefficients[,1:2], 3 )
```

```
##             Estimate Std. Error
## (Intercept)   -3.102      0.219
## age           -0.176      0.068
## smoke          0.399      0.273
```

Which of the fixed effects is significant?

Model Comparison

Similar to GLMs, we can compare models using the AIC.

How is the AIC defined?

Model Comparison

Similar to GLMs, we can compare models using the AIC.

How is the AIC defined?

Let's consider the seeds data set.

If we assume $p_1 = \ldots, p_{10}$, the AIC is

```
## [1] 64.30628
```

For the GLMM, the AIC is

```
## [1] 25.1
```

Which model would we choose?

The GLMM because it has the lower AIC.