Subtleties in mixed effects models

Let's talk **linear** mixed models

- Here is a key result from the **Gauss–Markov theorem**, which I mentioned earlier this semester:
  - If errors in a linear regression model are **uncorrelated** with **equal variance** and **expectation zero**, then ordinary least squares (OLS) is the **best** (minimum variance) **linear unbiased estimator** (BLUE) of the model coefficients β
- In the most basic linear regression setting – IID normal errors – this assumption is met:

$$\mathbf{y} = \beta\mathbf{X} + \epsilon,\ \ \epsilon \sim N(0,\sigma I),\ I \text{ being the identity matrix}$$

- What about a mixed linear model?

$$y_{ij} = \beta X_{ij} + b_i Z_{ij} + \epsilon_{ij},\ \ b \sim N(0,\Sigma_b),\ \ \epsilon \sim N(0,\sigma_\epsilon)$$

Correlated errors in linear mixed models

$$y_{ij} = \beta X_{ij} + b_i Z_{ij} + \epsilon_{ij}, \;\; b \sim N(0, \Sigma_b), \;\; \epsilon \sim N(0, \sigma_\epsilon)$$

- The stochastic part of y_{ij} is a linear combination of jointly Gaussian random variables (the b_is and ϵ_{ij}) with weights given by the vector Z_{ij}
- Recall from early in the semester: a linear combination of Gaussian random variables is also Gaussian!
  - And the mean is the weighted sum of the means!
- This means that a linear mixed model with IID b's and ϵ's can be re-expressed as a **simple linear regression**...

$$y_{ij} = \beta X_{ij} + \epsilon_{ij}$$

- ...with **correlated intra-group Gaussian errors** $\{\epsilon_{i\cdot}\}$

Correlated errors in linear mixed models

$$y_{ij} = \beta X_{ij} + \epsilon_{ij} \text{ with correlated mean-zero errors } \{\epsilon_{ij}\}$$

- An interesting (and possibly puzzling) consequence:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y \qquad \text{(OLS regression equation)}$$

$$E[\hat{\beta}] = (X^\top X)^{-1} X^\top E[Y] \qquad \text{(linearity of the expectation)}$$

$$= (X^\top X)^{-1} X^\top X \beta \qquad \text{(errors are mean-zero)}$$

$$= \beta \qquad \rightarrow \text{OLS is unbiased!!!}$$

- We'll now look at OLS versus mixed-linear model estimation of a simple simulated example where the true generative process is a mixed linear model

Much ado about nothing: **singular fits**

```
Linear mixed model fit by REML ['lmerMod']
Formula: value ~ 1 + condition + (1 | participant)
   Data: df.slopes

REML criterion at convergence: 83.6

Scaled residuals:
    Min      1Q  Median      3Q     Max 
-3.5808 -0.3184  0.0130  0.4551  2.0913 

Random effects:
 Groups      Name        Variance Std.Dev.
 participant (Intercept) 0.0000   0.0000  
 Residual                0.4512   0.6717  
Number of obs: 40, groups:  participant, 20

Fixed effects:
             Estimate Std. Error t value
(Intercept)  0.190524   0.150197   1.268
condition2  -0.001941   0.212411  -0.009

Correlation of Fixed Effects:
           (Intr)
condition2 -0.707
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Note how the summary says “singular fit”, and how the variance for random intercepts is 0. Here, fitting random intercepts did not help the model fit at all, so the lmer gave up …

https://psych252.github.io/psych252book/linear-mixed-effects-models-2.html

Much ado about nothing: **singular fits**

"A common issue with "maximal" linear mixed-effects models is that not all random effects can be estimated. These estimation issues are noted by warnings from lme4 such as boundary (singular) fit or Model failed to converge. These warnings indicate that the model is too complex for the data at hand."

https://mspeekenbrink.github.io/sdam-r-companion/linear-mixed-effects-models.html

Much ado about nothing: **singular fits**

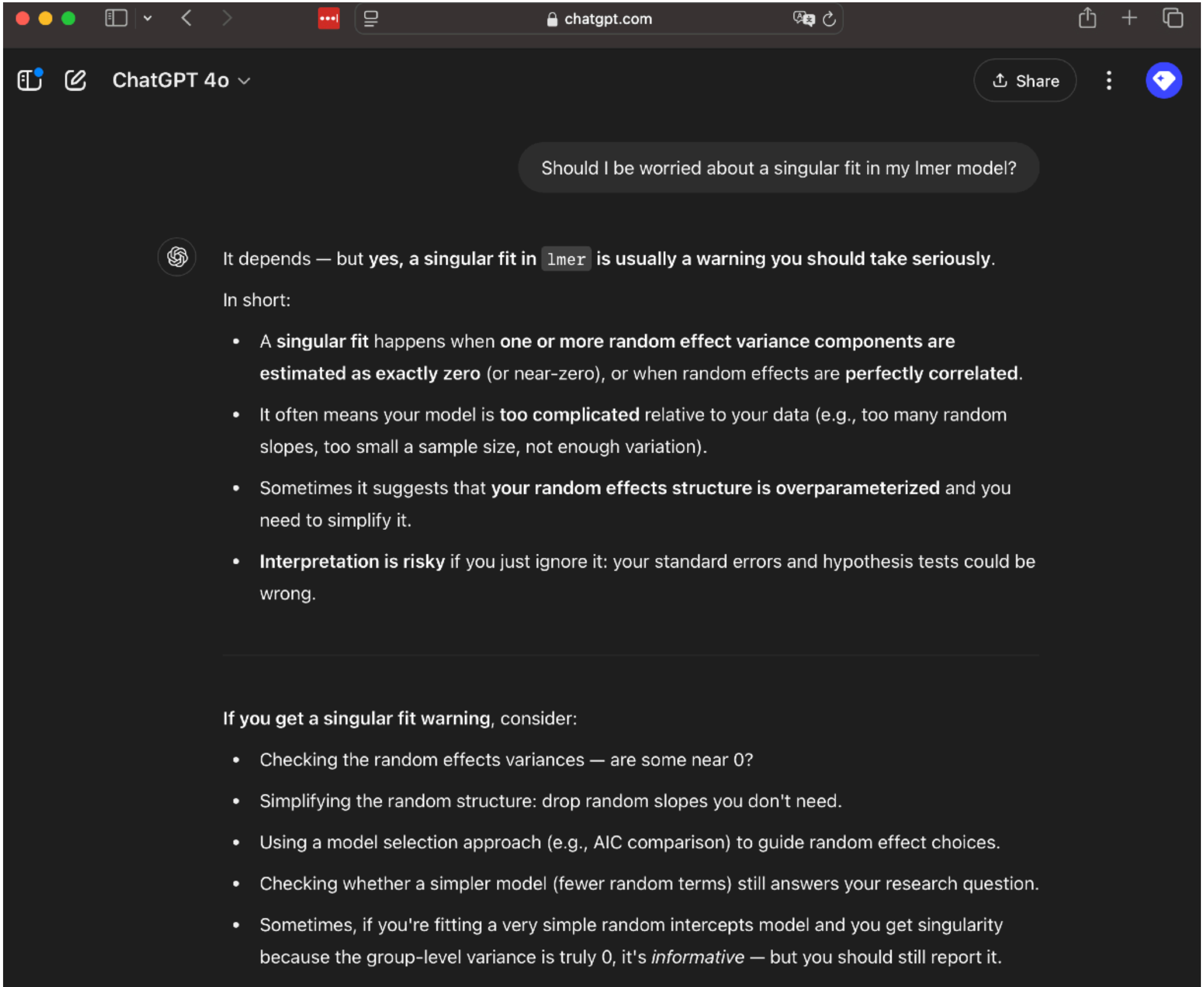
It is very common for overfitted mixed models to result in singular fits. Technically, singularity means that the random effects variance-covariance matrix is of *less than full rank*.
[...]
At present there are a variety of strong opinions about how to resolve such problems, which are sometimes conflated with the general problem of how to decide on the appropriate complexity of the random-effects component of a model. Briefly:
[...]

- Barr et al. (2013) suggest always starting with the maximal model (i.e. the most random-effects component of the model that is *theoretically* identifiable given the experimental design) and then dropping terms when singularity or non-convergence occurs (please see the paper for detailed recommendations …)

*The word **singular** does not appear in Barr et al. 2013!!!*

https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Much ado about nothing: **singular fits**



Should I be worried about a singular fit in my lmer model?

It depends — but **yes, a singular fit in `lmer` is usually a warning you should take seriously.**

In short:

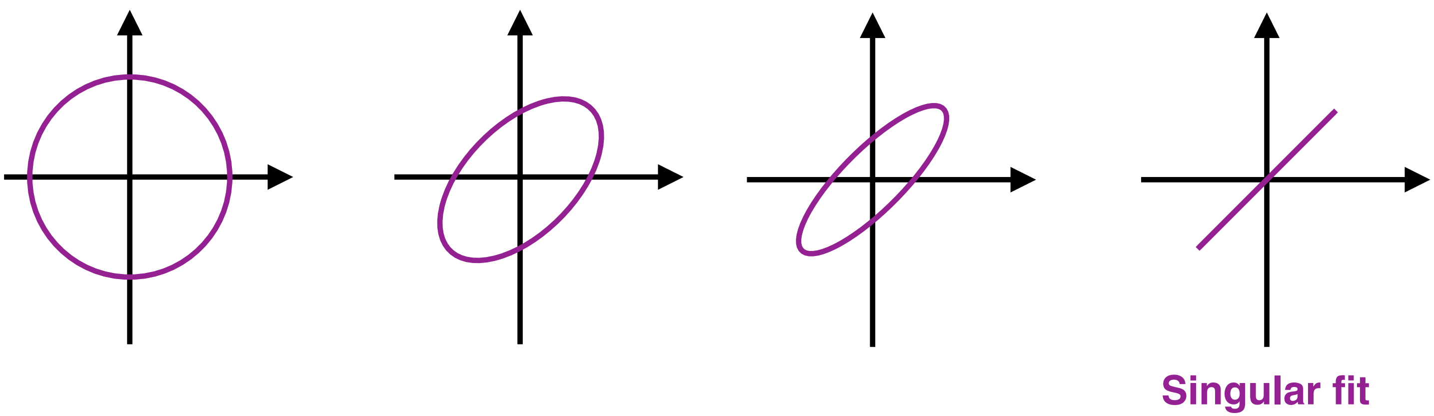
- A **singular fit** happens when **one or more random effect variance components are estimated as exactly zero** (or near-zero), or when random effects are **perfectly correlated**.
- It often means your model is **too complicated** relative to your data (e.g., too many random slopes, too small a sample size, not enough variation).
- Sometimes it suggests that **your random effects structure is overparameterized** and you need to simplify it.
- **Interpretation is risky** if you just ignore it: your standard errors and hypothesis tests could be wrong.

If you get a singular fit warning, consider:

- Checking the random effects variances — are some near 0?
- Simplifying the random structure: drop random slopes you don't need.
- Using a model selection approach (e.g., AIC comparison) to guide random effect choices.
- Checking whether a simpler model (fewer random terms) still answers your research question.
- Sometimes, if you're fitting a very simple random intercepts model and you get singularity because the group-level variance is truly 0, it's *informative* — but you should still report it.

The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0,\Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z



- This may well maximize the data likelihood – so **leave it alone!**

The generally correct thing to do about singular fits

The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0, \Sigma_b)$

The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0,\Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**

The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0, \Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z

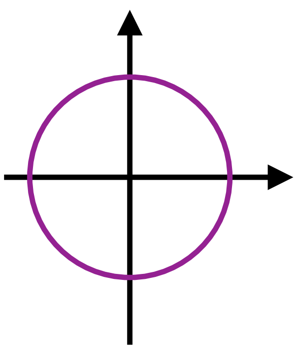
The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0,\Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z

- If this maximizes the data likelihood, then it maximizes the data likelihood – so **leave it alone!**

The generally correct thing to do about singular fits

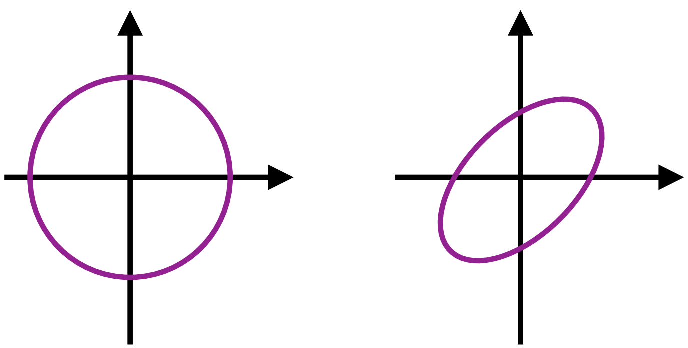
- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0, \Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z



- If this maximizes the data likelihood, then it maximizes the data likelihood – so **leave it alone!**

The generally correct thing to do about singular fits

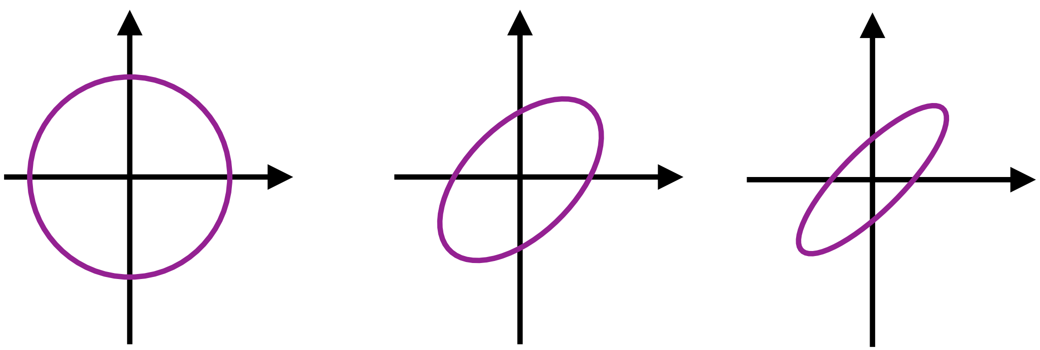
- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0, \Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z



- If this maximizes the data likelihood, then it maximizes the data likelihood – so **leave it alone!**

The generally correct thing to do about singular fits

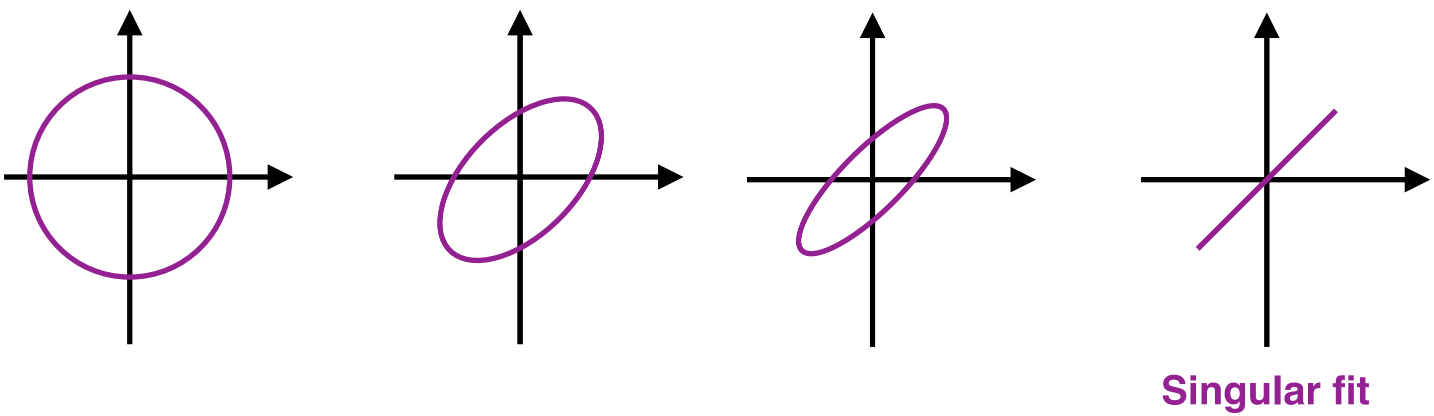
- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0,\Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the the vector space defined by Z



- If this maximizes the data likelihood, then it maximizes the data likelihood – so **leave it alone!**

The generally correct thing to do about singular fits

- Remember, a mixed effects model, the random effects b are assumed to be drawn from a multivariate normal distribution: $b \sim N(0, \Sigma_b)$
- A "singular fit" simply means that the random-effects covariance matrix Σ_b is **less than full rank**
- That is, the distribution of b lives on a **subspace** of the vector space defined by Z



- If this maximizes the data likelihood, then it maximizes the data likelihood – so **leave it alone!**