

Bayesian analyses for bilingualism research

ISB14 Workshop

Introduction to Advanced Statistics for Language Sciences

João Veríssimo (University of Lisbon)

2023-06-23

Packages

- Accompanying script/data: <https://tinyurl.com/BayesianISB14>
- R script: `verissimo_ISB14.R`
- Packages required to run the accompanying script:

```
library(lme4)  
library(brms)
```

```
library(ggplot2)  
library(ggeffects)  
library(ggdist)
```

Background

Bayesian vs. frequentist

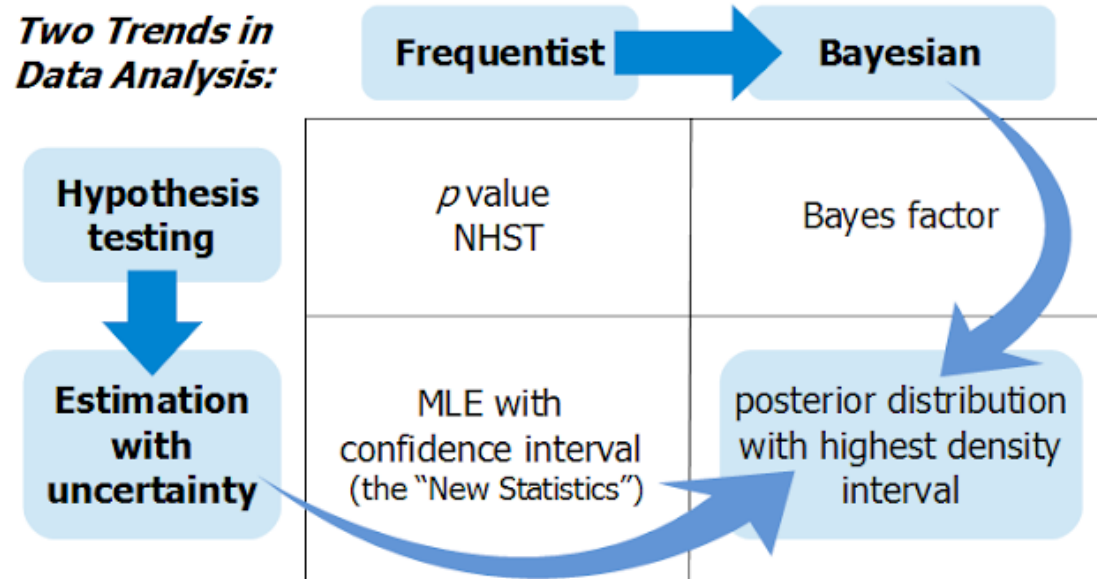
- Bayesian approaches are usually presented in opposition to frequentist ones
- In statistics proper, there is a *debate* about the two frameworks (about substantial, even philosophical, issues)

Bayesian vs. frequentist

- This session takes a more pragmatic stance
- Bayesian statistics often allow us to do *more things*
- And to do them within an integrated (conceptual and technical) framework

Two trends in data analysis

Kruschke (2018)



Copyright © 2015 John K. Kruschke

NHST: Null Hypothesis Significance Testing

MLE: Maximum Likelihood Estimation

"The New Statistics": Book/paper by Geoff Cumming (2011, 2014)

Bayesian vs. frequentist

- The goal of this session is not to 'convert' people
- There are certainly many reasonable applications of hypothesis testing and of frequentist statistics
- The goal is to introduce you to a *framework* and associated *tools* ...
- ... which I believe have substantial **advantages** over the more traditional statistical approaches

Some advantages of Bayesian statistics

- Solves (or at least, ameliorates) some long-standing problems with NHST
- Facilitates a focus on estimation and uncertainty when assessing effects
- Provides more conceptually sound statistical inference
- Allows bringing in prior information into the modelling
- Facilitates modelling of variance components
- Greater flexibility: every quantity can be directly compared to any other
- Accommodates many different distribution families
- Robustness to noise and extreme values
- Allows assessing evidence for null values

Problems with NHST

- Despite the wide adoption of frequentist statistics, it is known to have important problems^{*}

^{*} Or better said: NHST and the way it is employed is known to have important problems

Problems with NHST

- A first problem with NHST is its emphasis on binary decisions
- Binary decisions ('effect' vs. 'no effect') ...
 - Impoverish our statistical inferences
 - Encourage the application of mechanical, mindless, and dogmatic procedures

Problems with NHST

- “The dogmatism has lasted for almost half a century. This is far too long. We need a knowledgeable use of statistics, not a collective compulsive obsession.” (Gigerenzer, 1993)
- “Do not replace the dogmatism by a new, altogether different one (e.g., Bayesian dogmatism).” (Gigerenzer, 1993)

Fundamental Bayesian principles

Three Bayesian principles

1. Parameters as full probability distributions
2. Bayesian posteriors give us the probability of the different values of a parameter, given the data
3. Beliefs are updated through the integration of prior knowledge

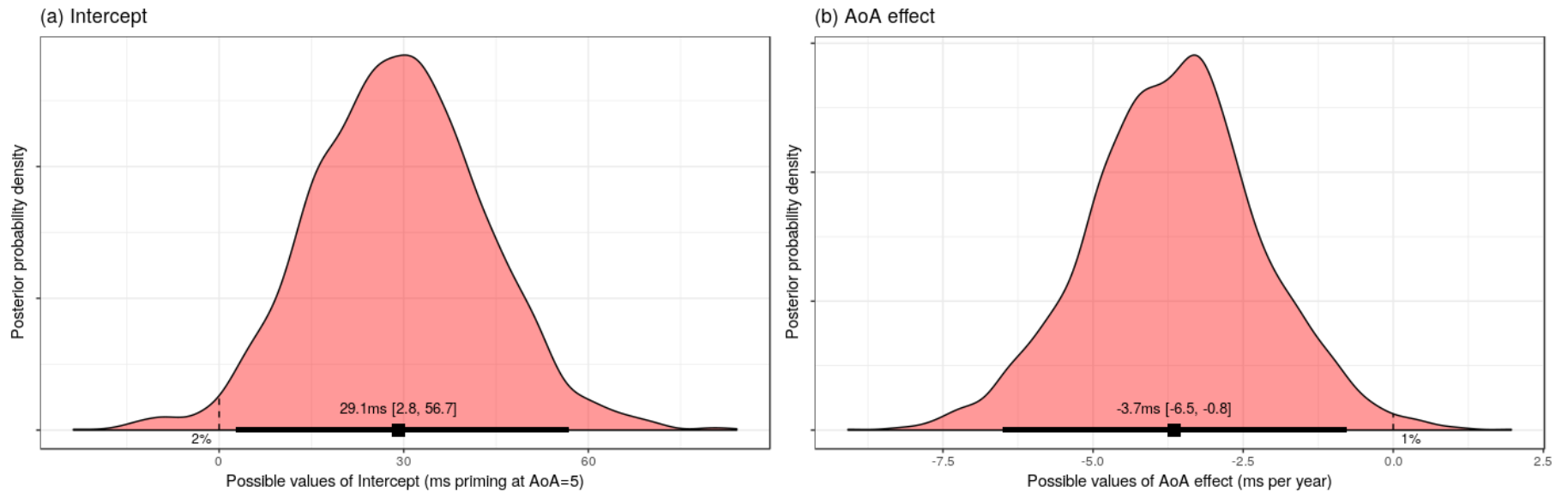
Three Bayesian principles

1. **Parameters as full probability distributions**
2. Bayesian posteriors give us the probability of the different values of a parameter, given the data
3. Beliefs are updated through the integration of prior knowledge

Parameters as full distributions

- Bayesian estimates are not single fixed values
(e.g., “the regression slope of the difference between means is 70ms”)
- Estimates of parameters are *full probability distributions* over a range of values

Parameters as full distributions



Examples of posterior distributions: (a) regression intercept and (b) slope.

Your first Bayesian model

- The data for this example comes from a study with L2 learners in which morphological priming was predicted by age of onset of L2 acquisition (AoA) (Veríssimo et al., 2018, *Lang Acq*)
- Participants were presented with 50ms subliminal primes that were morphologically related to a target word (to which a lexical decision was made)
- There were three types of prime words (in German, actually):
 - Inflectional: e.g., *boiled-BOIL*
 - Derivational: e.g., *boiler-BOIL*
 - Unrelated: e.g., *parked-BOIL*

Load the data

```
priming.aoa <- readRDS("verissimo2018_aoa5.rds")  
head(priming.aoa)
```

```
##      Subject Inf.Priming AoA.Ger  
## 33      501    -105.819      5  
## 34      502     104.011      5  
## 35      503     -17.621      6  
## 36      504      30.501      5  
## 37      505     146.580      5  
## 39      507      28.804      5
```

Dataset

- To simplify, data has been reduced and aggregated:
- This is a subset of 47 subjects (those with AoA>5)
(the paper reported a more complex non-linear pattern on a larger dataset)
- By-participant means and priming effects were calculated:
 $\text{Inf.Priming} = \text{RT}(\text{Unrelated}) - \text{RT}(\text{Inflectional})$
- Aggregation was done on logged RTs and back-transformed to ms
(to yield a more-or-less normal distribution)
- We'll perform a simple regression, no covariates

Frequentist and Bayesian models

1. We will first run a **frequentist** linear regression, in which we predict inflectional priming from AoA
2. We will then run a **Bayesian** linear regression model

Frequentist model

Linear regression

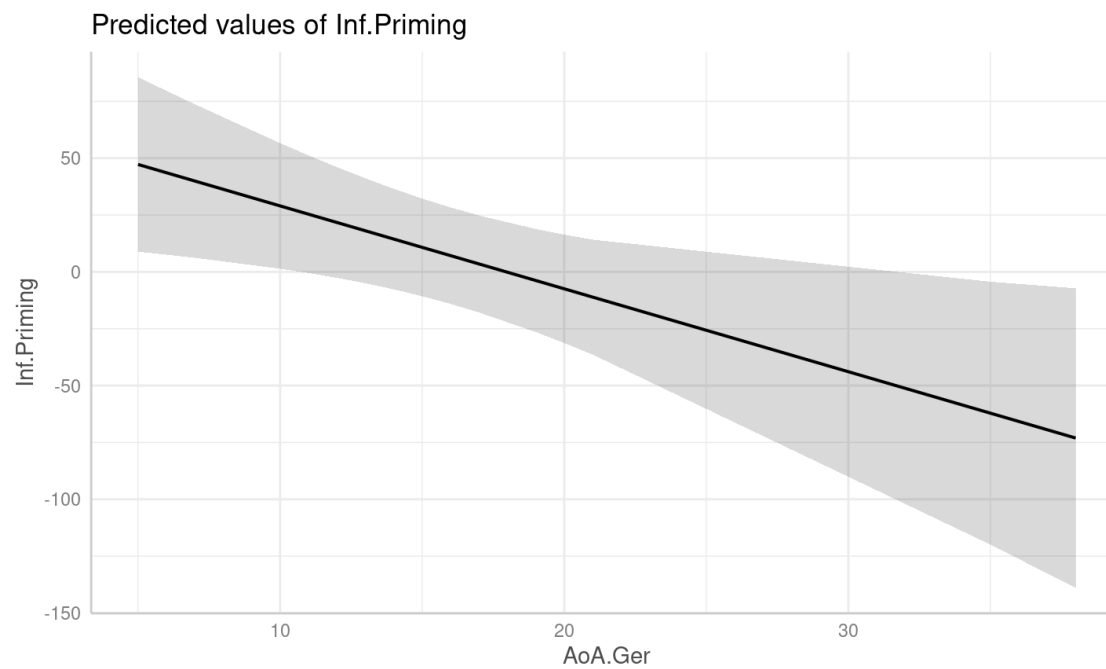
```
c. <- function(x) scale(x, scale=F) # Centering function
summary(m.freq <- lm(Inf.Priming ~ c.(AoA.Ger), priming.aoa)) # Model
```

```
##
## Call:
## lm(formula = Inf.Priming ~ c.(AoA.Ger), data = priming.aoa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -164.83  -30.53   3.08   39.46  235.79
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.45      10.52   0.61  0.543
## c.(AoA.Ger)    -3.65       1.42  -2.57  0.014 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72.1 on 45 degrees of freedom
## Multiple R-squared:  0.128, Adjusted R-squared:  0.108
## F-statistic: 6.6 on 1 and 45 DF, p-value: 0.0136
```

Frequentist model

Plot of fitted values

```
ggemmeans(m.freq, terms = "AoA.Ger") |> plot()
```



Bayesian model

- This will be very simple to set up
 - We will use the *brms* package (Bürkner, 2017)
 - Its function `brm()` allows fitting Bayesian regression models
 - It takes the same kind of formulas as `lm()` or `lmer()`
- However, one may run into memory problems or other difficulties...
(the model has already been fitted and can be loaded and inspected)
- Also: A *real* Bayesian analysis would require a few more steps!

Bayesian model

```
summary(m.bay <- brm(Inf.Priming ~ c.(AoA.Ger),  
                    data = priming.aoa))
```

```
## Family: gaussian  
## Links: mu = identity; sigma = identity  
## Formula: Inf.Priming ~ c.(AoA.Ger)  
## Data: priming.aoa (Number of observations: 47)  
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup draws = 4000  
##  
## Population-Level Effects:  
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## Intercept      6.55     10.46  -14.40   26.33 1.00     3362     2625  
## c.AoA.Ger     -3.63      1.43   -6.48   -0.85 1.00     3750     2711  
##  
## Family Specific Parameters:  
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## sigma      72.67      7.89   59.35   90.24 1.00     3317     2828  
##  
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS  
## and Tail_ESS are effective sample size measures, and Rhat is the potential  
## scale reduction factor on split chains (at convergence, Rhat = 1).
```


Parameters as full distributions

- The reported estimate is just the mean of the full probability distribution
- The distribution is made up of *samples* that are collected during the fitting process (4,000 samples, in this case)
- We can obtain them as such (note that fixed effects are prefixed with 'b_'):

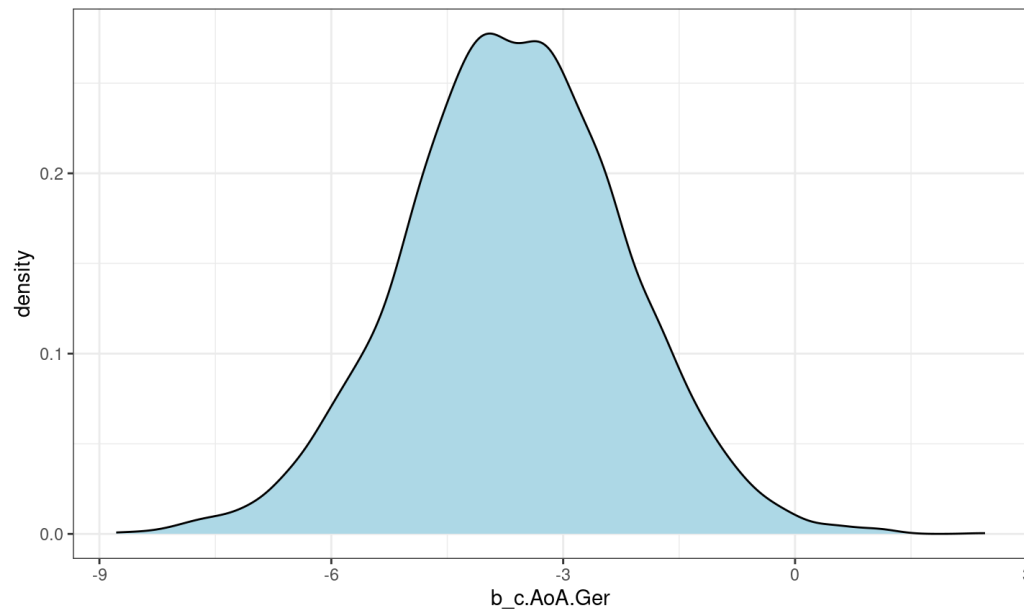
```
(ps <- as_draws_df(m.bay))           # Posterior samples for every parameter
```

```
## # A draws_df: 1000 iterations, 4 chains, and 5 variables
##   b_Intercept b_c.AoA.Ger sigma lprior lp__
## 1      20.8      -4.0     58   -9.9 -276
## 2      17.6      -2.5     63  -10.0 -274
## 3      16.0      -2.1     70  -10.1 -274
## 4      16.1      -1.8     72  -10.2 -274
## 5       8.9      -4.9     72  -10.2 -273
## 6      16.2      -3.2     72  -10.2 -273
## 7     -12.2      -4.6     69  -10.2 -275
## 8     -12.5      -6.2     70  -10.2 -276
## 9     -11.5      -6.0     83  -10.5 -276
## 10      3.6      -4.2     68  -10.1 -273
## # ... with 3990 more draws
## # ... hidden reserved variables {'.chain', '.iteration', '.draw'}
```

Posterior distributions

- This is the *posterior distribution* of the parameter
- It is very informative about the possible values of parameters:

```
ggplot(ps, aes(x=b_c.AoA.Ger)) + theme_bw() +  
  geom_density(fill="lightblue")
```



Posterior distributions, estimation, and uncertainty

Three Bayesian principles

1. Parameters as full probability distributions
2. **Bayesian posteriors give us the probability of the different values of a parameter, given the data**
3. Beliefs are updated through the integration of prior knowledge

Another problem with NHST

- The p -value is often incorrectly interpreted
- Often interpreted as the probability that H_0 is true
- Or as the probability that a (rather unspecified) alternative hypothesis is true
(= $1 - p$)
- p -values do not provide such probabilities

Two types of conditional probability

- These two questions are very different:
 - “Given that H_0 is true, what is the probability of obtaining this data?”
 - “Given that we have obtained this data, what is the probability of H_0 ?”
- Our puny human brains have trouble with this distinction (even if we’re scientists)

So what are p -values?

- A p -value is the probability of obtaining this data (or more extreme), assuming that *the null hypothesis is true*:

$$P(D|H_0)$$

- This is quite different from what we would like to know: the probability of certain hypotheses or parameter values, *given the obtained data*:

$$P(H_0|D)$$

The 'Bayes' in 'Bayesian'



Bayes' theorem

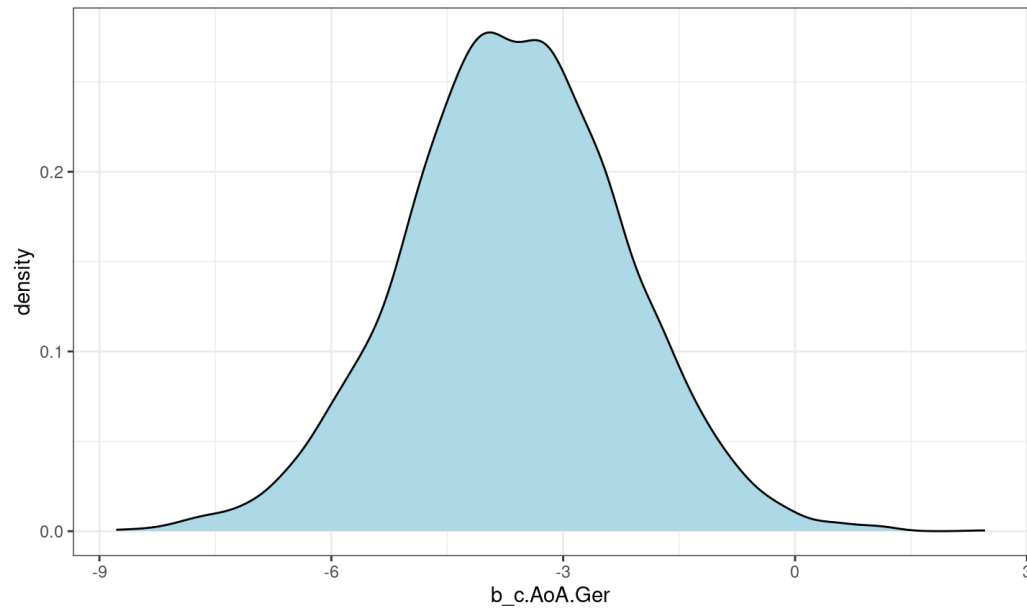
Thomas Bayes (1763)

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Posterior distributions

As probabilities of parameter values

```
ggplot(ps, aes(x=b_c.AoA.Ger)) + theme_bw() +  
  geom_density(fill="lightblue")
```



Posterior distributions and uncertainty

- Such posterior probability distributions formalise our *uncertainty*!
- If they are very wide, many values are possible (we simply do not know)
- If they are very narrow, we can be confident that the 'true' value is inside a small region of possible values

Credible intervals

- Bayesian statistics provides *credible intervals*, not confidence intervals
- They can be easily obtained as such:

```
posterior_summary(ps$b_c.AoA.Ger)
```

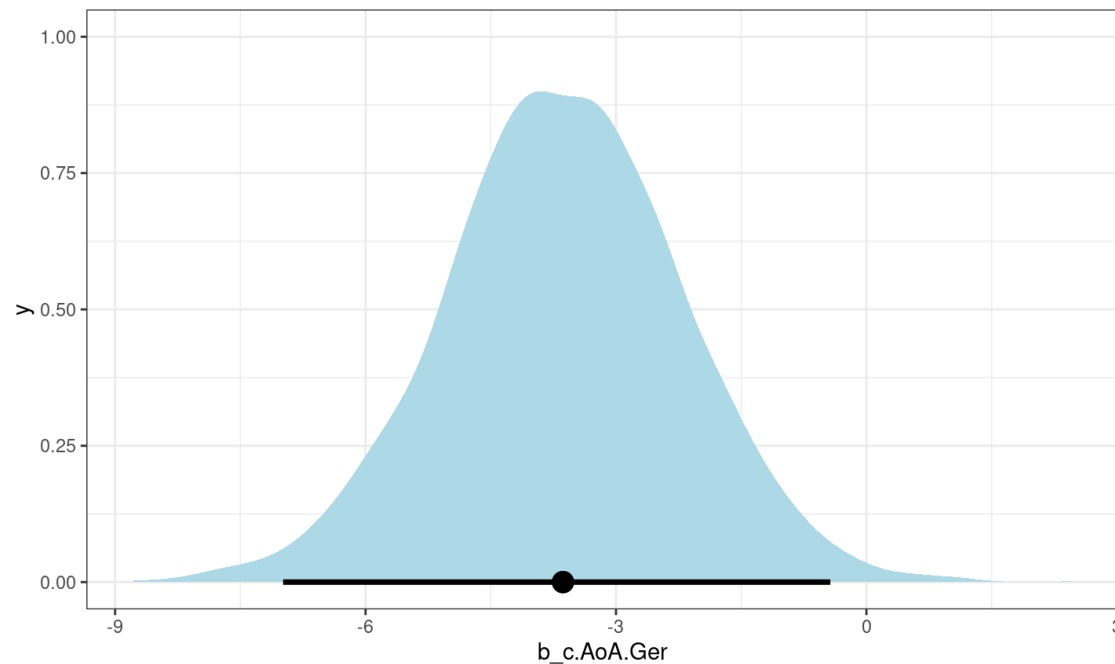
```
##      Estimate Est.Error   Q2.5   Q97.5  
## [1,]   -3.635    1.4287 -6.483  -0.8467
```

- They can be reported as such:
“an effect of AoA of -3.63ms per year, 95% CrI [-6.48, -0.85]”

Credible intervals

- The *ggdist* package is particularly good for visualising posterior distributions and their *credible intervals*:

```
ggplot(ps, aes(x=b_c.AoA.Ger)) + theme_bw() +  
  stat_halfeye(point_interval = mean_qi, .width = c(.025, .975),  
              fill="lightblue", interval_color="black", interval_size=5, point_size=4)
```



Credible intervals

- There is a 95% probability that the true value (of a mean, a slope, etc.) is inside the 95% credible interval
- This is usually how confidence intervals are interpreted, but wrongly so (because frequentist statistics does not allow computing probabilities of parameter values)

So ... is there an effect?

- From the point of view of estimation, this is not the right question
- Asking whether “there is an effect” reduces such statements about magnitudes and probabilities to a simple *directional* statement: AoA effect < 0
- Instead (or additionally), we are interested in quantifying the probabilities of different magnitudes of the AoA effect on inflectional priming

Emphasis on estimation

- The most probable effect of AoA is a reduction of inflectional priming by 3.6ms per year, with a 95% probability that this reduction is within -6.48 and -0.85.
- The most probable values for the slope of AoA are negative; thus, it is very likely that each year of AoA reduces inflectional priming by a few ms.
- The magnitude of this effect is likely to be relatively large: a person who learned German 10 years later than another is expected to show a substantial reduction in inflectional priming (which is about 30–40ms in L1):

```
posterior_summary(ps$b_c.AoA.Ger * 10)
```

```
##      Estimate Est.Error   Q2.5   Q97.5  
## [1,]   -36.35    14.287  -64.83  -8.467
```

- At the same time, it is possible (but unlikely) that the true effect will turn out to be quite small (<0.8ms/year, with ~2.5% probability)

Bringing in prior knowledge

Three Bayesian principles

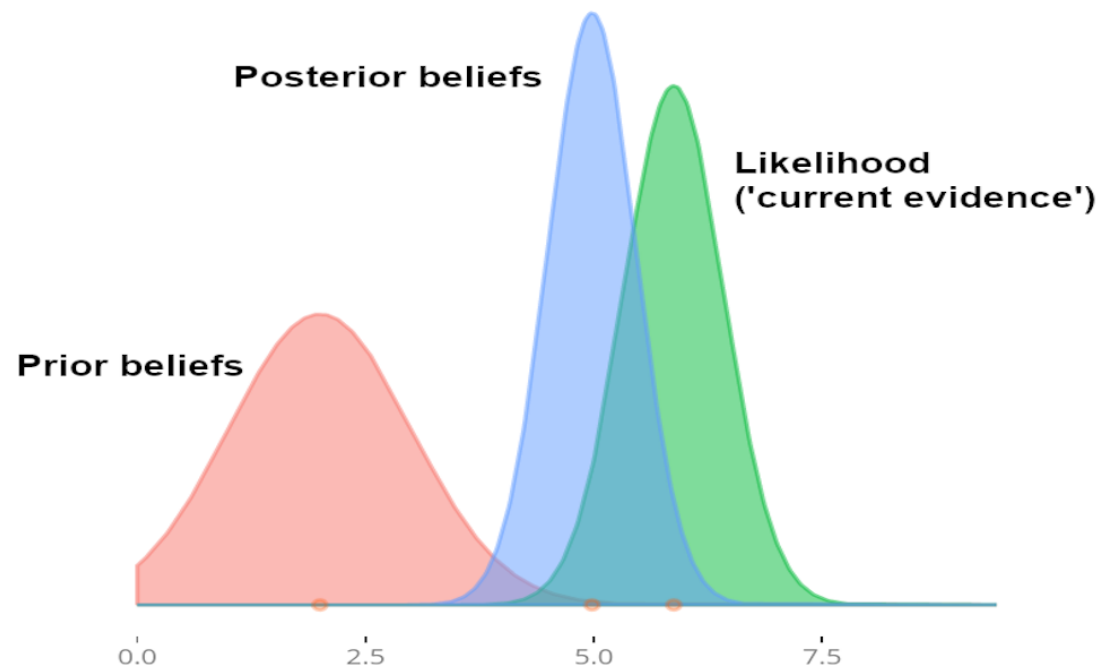
1. Parameters as full probability distributions
2. Bayesian 'posteriors' give us the probability of the different values of a parameter, given the data
3. **Belief updating through the integration of prior knowledge**

Prior knowledge

- Determining the probability of a hypothesis or of a parameter value from the data (e.g., that $A - B = 20ms$) ...
- ... depends critically on how likely this estimate is *a priori*, independently of the data

Priors, likelihood, posteriors

- Just as posterior beliefs are full probability distributions, so are *prior beliefs*
- And so is the 'current evidence', also called the *likelihood*



Prior knowledge

We already use it!

- How much would you believe the following (made-up) claims?
 - IQ is lower in bilinguals, compared to monolinguals ($p < .05$)
 - Risk of dementia is halved in bilinguals, compared to monolinguals ($p < .05$)
 - There is no difference in reading times between congruent and incongruent words in the color Stroop task ($p > .05$, n.s.)

Prior knowledge

We already use it!

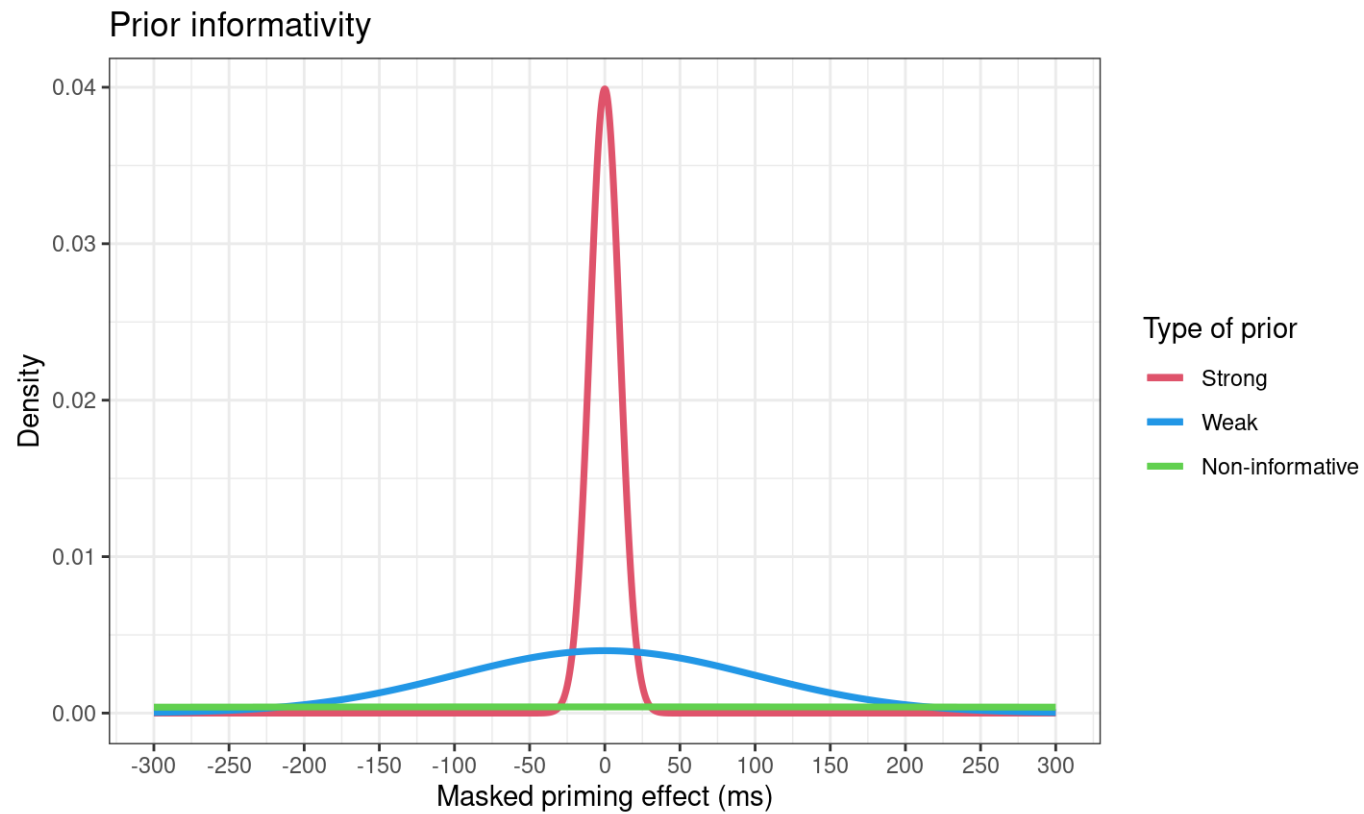
- Outcomes that are obvious or expected (because they are more consistent with previous evidence) are more believable as general claims
- Outcomes that are surprising, unexpected, or that clearly challenge previous evidence are less believable as general claims

Prior knowledge

We already use it!

- We already bring our prior knowledge into play when inferring from results!
- Bayesian statistics requires us to formalise our prior knowledge
- ... and integrates it into the statistical analysis

Informativity of priors



Note: A truly informative prior might be centered elsewhere, rather than at 0

Priors, likelihood, posteriors

- Strong prior, weak data
 - The influence of the prior is much larger when the likelihood is weak (i.e., when the data is very variable or from small samples)
 - In such cases, the prior 'compensates' for the noise in the data

Priors, likelihood, posteriors

- Weak prior, strong data
 - The influence of the likelihood is much larger when the prior is 'weak'
 - In such cases, our previous beliefs are not very constraining
 - We admit many possibilities and let the data 'speak for itself'
- **Extraordinary claims require extraordinary evidence**

Which priors are available?

- The `prior_summary()` function extracts priors from fitted models:

```
prior_summary(m.bay)
```

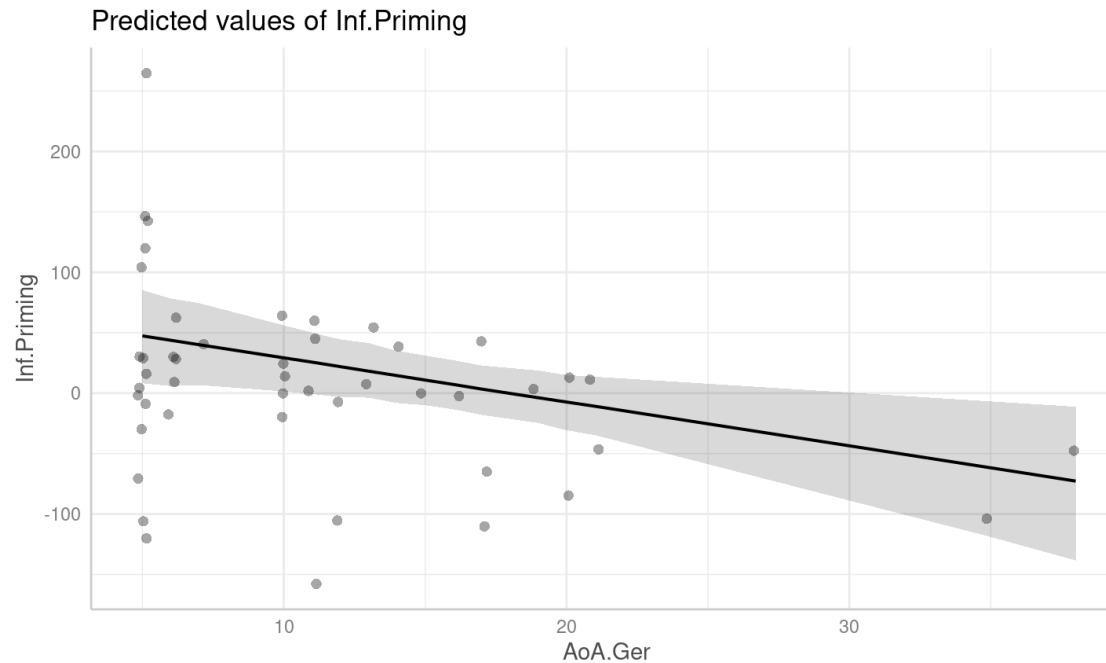
```
##           prior      class      coef group resp dpar nlpar lb ub
##           (flat)         b
##           (flat)         b c.AoA.Ger
## student_t(3, 7.4, 49) Intercept
## student_t(3, 0, 49)   sigma
## source
## default
## (vectorized)
## default
## default
```

- We can see that flat priors (i.e., non-informative) were used for the slope

Could priors help our inference?

- The effect of AoA on priming is quite large, perhaps implausibly so

```
ggemmeans(m.bay, terms = "AoA.Ger") |> plot(add.data=T)
```



- In particular, the model predicts large inhibitory effects, which are not likely from a 50ms short-duration prime

Defining priors

- Perhaps our inferences can be improved by using a more constraining prior ...
- The response variable (inflectional priming) is in ms, so priors should express the effects that we would expect to see in ms units (per year of AoA, in the case of the slope)

Model with stronger priors

- We start by defining our stronger prior on the regression slope:

```
(priors.strong <- c(prior(normal(0, 250), class=Intercept),  
  prior(normal(0, 2.5), class=b),  
  prior(normal(0, 250), class=sigma)))
```

```
##           prior      class coef group resp dpar nlpars lb  ub source  
## normal(0, 250) Intercept                <NA> <NA>  user  
## normal(0, 2.5)         b                  <NA> <NA>  user  
## normal(0, 250)        sigma                <NA> <NA>  user
```

- And fit the new model:

```
m.bay.strong <- brm(Inf.Priming ~ c.(AoA.Ger), priming.aoa,  
  prior = priors.strong,  
  sample_prior = "yes",  
  file = "m-bay-strong.rds")
```

Is the stronger prior justifiable?

- A Normal(0, 2.5) prior is a Normal distribution with an SD of 2.5
- We know what the properties of such a distribution are:
on a Normal, 95% of the mass is between the mean plus or minus 2 SDs
- So we are effectively saying that it is quite likely, *a priori* (with 95% probability) that the effect of AoA on inflectional priming is of up to 5ms per year
- Given the documented differences between L1 speakers and late learners, this seems appropriate

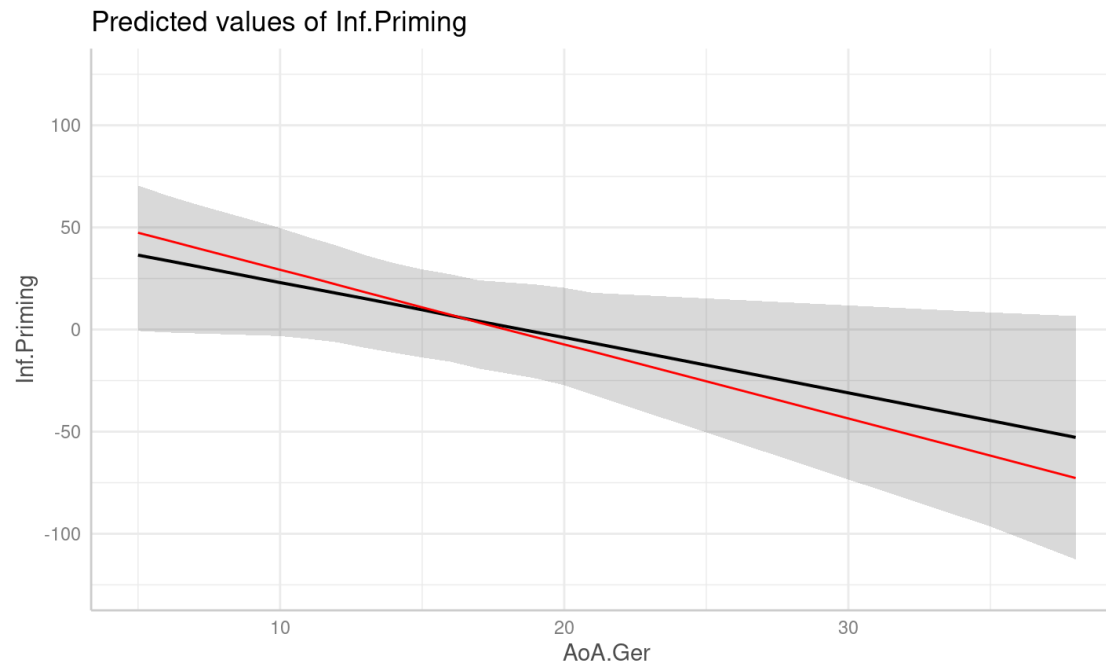
Model with strong priors

```
fixef(m.bay.strong)
```

##	Estimate	Est.Error	Q2.5	Q97.5
## Intercept	6.3428	10.8218	-15.4992	27.84784
## c.AoA.Ger	-2.6936	1.2878	-5.2012	-0.14203

Regression line (w/ strong priors)

```
line.flat <- ggemmeans(m.bay, terms = "AoA.Ger")  
ggemmeans(m.bay.strong, terms = "AoA.Ger") |> plot() +  
  scale_y_continuous(limits = c(-125, 125)) +  
  geom_line(data=line.flat, aes(y=predicted, x=x), col="red")
```



How do we choose priors?

- General knowledge about sizes of effects in our field
- Specific knowledge about task, measure, population, ...
- Meta-analyses of similar effects
- Effects from similar studies or pilot data
- Other general guidelines:
<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>

Do priors 'influence' the interpretation?

- Yes... and there's nothing wrong with that!
- We do possess such prior knowledge and it is already implicitly used when we interpret results
- Priors are defined transparently and can be reasonably discussed
- They should be justified, like any other aspect of data analysis
- Remember that the goal is to arrive at the best and most generalizable inference... not the best inference for this sample!

Some advantages of Bayesian statistics

- Solves (or at least, ameliorates) some long-standing problems with NHST
- Facilitates a focus on estimation and uncertainty when assessing effects
- Provides more conceptually sound statistical inference
- Allows bringing in prior information into the modelling
- Facilitates modelling of variance components
- Greater flexibility: every quantity can be directly compared to any other
- Accommodates many different distribution families
- Robustness to noise and extreme values
- Allows assessing evidence for null values

Example of mixed-effects model

Load data

```
head(priming.lmm <- readRDS("verissimo_cunnings_inprep.rds"))
```

```
## Participant PrimeType Prime Target RT
## 1 1649318400 Inflected tuned TUNE 691
## 2 1649318400 Control mocked SKATE 760
## 3 1649318400 Inflected marked MARK 863
## 4 1649318400 Derived defender DEFEND 619
## 5 1649318400 Inflected lectured LECTURE 601
## 6 1649318400 Control chopped TOAST 634
```

Dataset

- This is a different dataset, but again, the data comes from a morphological priming study with L2 learners (conducted w/ Ian Cunnings)
- Participants were presented with overt visual primes (250ms) that were morphologically related to a target word (to which a lexical decision was made)
- There were again three types of prime words (in English):
 - Inflectional: e.g., *boiled-BOIL*
 - Derivational: e.g., *boiler-BOIL*
 - Unrelated: e.g., *parked-BOIL*

Number of responses by participant

```
with(priming.lmm, table(Participant, PrimeType)) |> head(10)
```

```
##           PrimeType
## Participant Control Derived Inflected
## 1649318400      26      32      32
## 1649318423      26      33      30
## 1649318445      29      29      30
## 1649318493      32      34      33
## 1649318535      22      32      27
## 1649318542      31      30      31
## 1649318589      25      30      31
## 1649318594      17      25      26
## 1649318742      34      33      34
## 1649318832      33      34      33
```


Frequentist LMM

Fixed effects

- We predict response times (RTs) on the basis of PrimeType (unrelated, inflected, derived)

```
lmm.freq <- lmer(RT ~ (1 + PrimeType | Participant) + 1 + PrimeType, priming.lmm)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(lmm.freq)$coefficients
```

```
##           Estimate Std. Error t value
## (Intercept)    721.654    11.6714  61.831
## PrimeTypeDerived  -75.339     7.0454 -10.693
## PrimeTypeInflected -72.321     6.9154 -10.458
```

Frequentist LMM

Random effects

VarCorr(lmm.freq)

##	Groups	Name	Std.Dev.	Corr
##	Participant	(Intercept)	70.89	
##		PrimeTypeDerived	9.98	1.00
##		PrimeTypeInflected	4.43	1.00 1.00
##	Residual		180.45	

Bayesian LMM

Model on raw RTs

```
lmm.bay <- brm(RT ~ 1 + PrimeType + (1 + PrimeType | Participant), priming.lmm,  
  cores=4, # In parallel (1 chain per core)  
  file = "lmm-bay.rds")
```

Bayesian LMM

Fixed effects

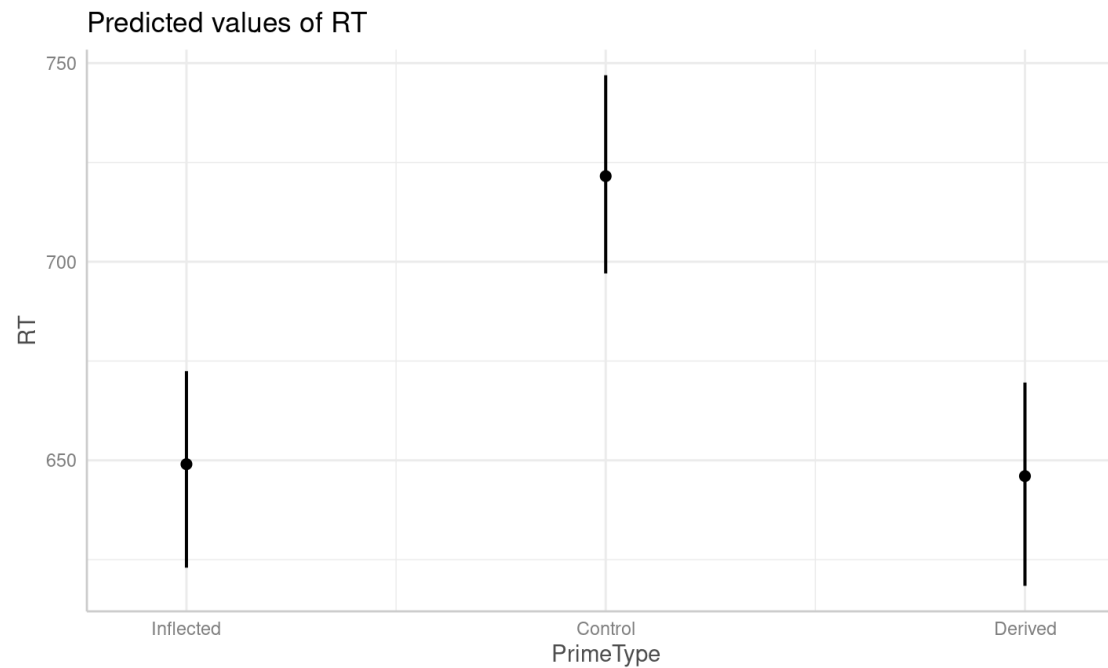
```
summary(lmm.bay)$fixed[1:4]
```

##	Estimate	Est.Error	l-95% CI	u-95% CI
## Intercept	721.380	12.5846	695.761	745.977
## PrimeTypeDerived	-75.692	6.8864	-88.994	-61.815
## PrimeTypeInflected	-72.571	7.0394	-86.284	-58.987

Bayesian LMM

Plot of fitted values

```
ggemmeans(lmm.bay, terms = "PrimeType") |> plot()
```



Bayesian LMM

Random effects

```
summary(lmm.bay)$random$Participant[1:4]
```

##	Estimate	Est.Error	l-95% CI	u-95% CI
## sd(Intercept)	77.221691	9.26050	61.40237	97.63482
## sd(PrimeTypeDerived)	9.361012	6.58128	0.47353	24.23401
## sd(PrimeTypeInflected)	9.850555	7.40936	0.34747	27.21504
## cor(Intercept,PrimeTypeDerived)	0.294424	0.44175	-0.70333	0.92874
## cor(Intercept,PrimeTypeInflected)	0.024114	0.44576	-0.78759	0.85871
## cor(PrimeTypeDerived,PrimeTypeInflected)	0.178683	0.49102	-0.79847	0.92439

Bayesian LMM

Residuals

```
summary(lmm.bay)$spec_pars[1:4]
```

```
##      Estimate Est.Error l-95% CI u-95% CI  
## sigma  180.48    2.0022  176.57  184.53
```

Estimating variance components

- One important advantage of Bayesian models is that one can better estimate the different variance components (i.e., the random effects)
- This is especially the case for the correlation parameters
- This means that more complex models can be fitted, with less errors of convergence (especially important when fitting maximal models)
- Moreover, all random effect parameters can be accompanied by credible intervals (in fact, by full posteriors)

Comparison of variance components

- We can even compare different variance components to answer questions like:

Is there more between-subject variance for one type of prime or the other?

Comparison of variance components

- We can even compare different variance components to answer questions like:

Is there more between-subject variance for one type of prime or the other?

- We simply subtract the posterior samples to obtain their comparison:

```
ps <- as_draws_df(lmm.bay)           # Posterior samples for every parameter
(ps$sd_Participant__PrimeTypeDerived - ps$sd_Participant__PrimeTypeInflected) |> # Subtraction of estimates
posterior_summary()
```

```
##      Estimate Est.Error   Q2.5  Q97.5
## [1,] -0.48954    8.8455 -18.43 16.529
```

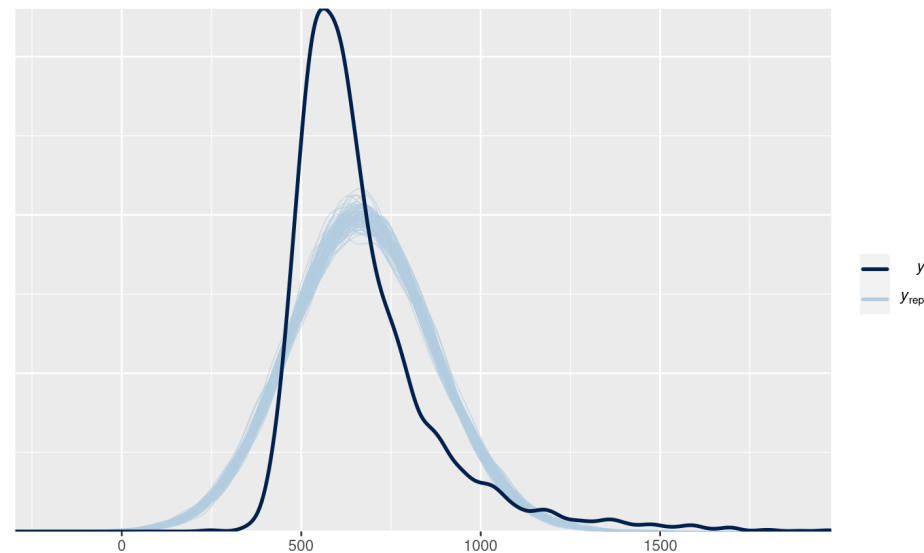
Some advantages of Bayesian statistics

- Solves (or at least, ameliorates) some long-standing problems with NHST
- Facilitates a focus on estimation and uncertainty when assessing effects
- Provides more conceptually sound statistical inference
- Allows bringing in prior information into the modelling
- Facilitates modelling of variance components
- Greater flexibility: every quantity can be directly compared to any other
- Accommodates many different distribution families
- Robustness to noise and extreme values
- Allows assessing evidence for null values

Posterior predictive checks

- One important model diagnostic in the Bayesian framework is the *posterior predictive check*
- We simulate data from the model and compare the simulated datasets to what we have obtained

```
pp_check(lmm.bay, ndraws = 100)
```



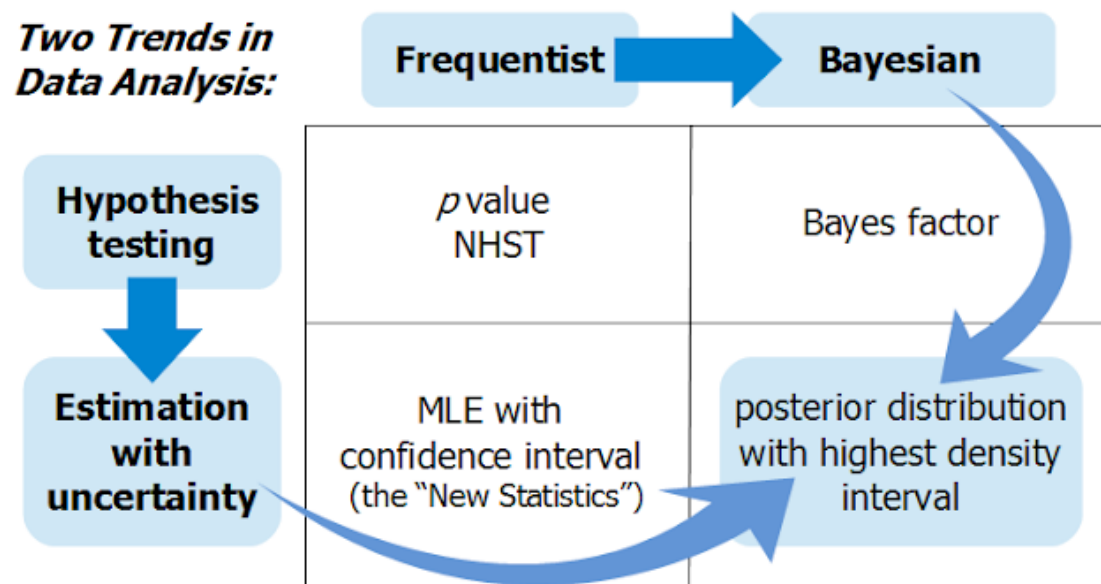
- This is a particularly bad model ... the reason is that RTs are not normally distributed

Many different distribution families

- RTs (or fixation times): Lognormal, shifted lognormal, ex-Gaussian ...
(see Ciaccio & Veríssimo, 2022, *PB&R*)
- Accuracy (or other nominal variables): Binomial, Bernoulli, ...
(see Lago et al., 2022, *Lang Learning*)
- Ratings (or other ordinal variables): Cumulative ordinal models
(see Veríssimo, 2022, *BLC*)
- Measures with many outliers: t distribution, ...
- RTs + accuracy: Wiener model, ...
- Fixation times + probability of fixation: Hurdle lognormal, ...

Two trends in data analysis

Kruschke (2018)



Copyright © 2015 John K. Kruschke

NHST: Null Hypothesis Significance Testing

MLE: Maximum Likelihood Estimation

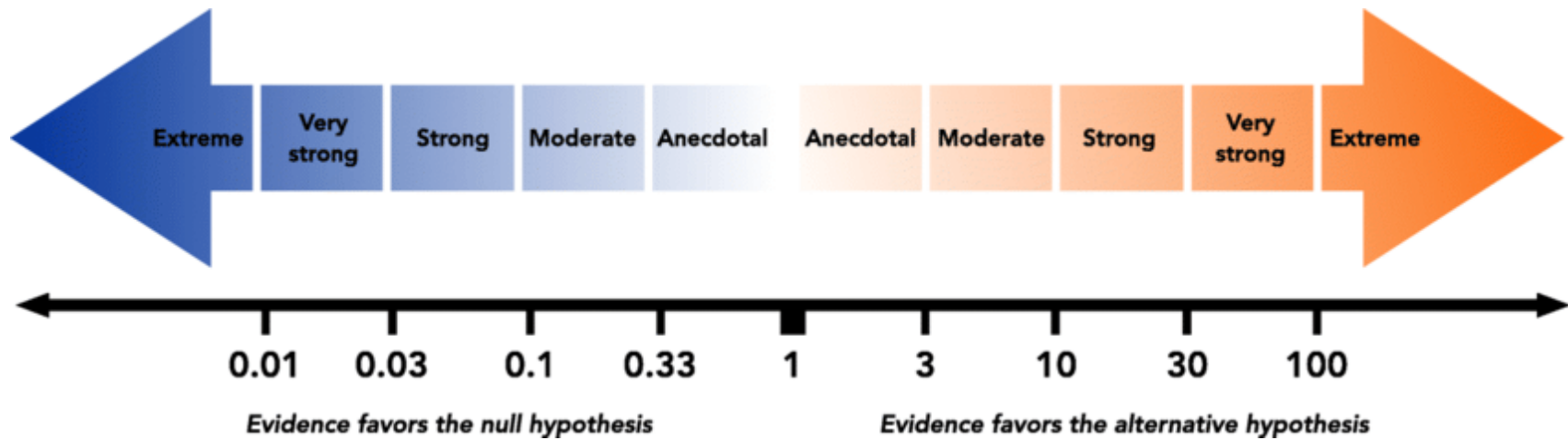
"The New Statistics": Book/paper by Geoff Cumming (2011, 2014)

Bayesian hypothesis testing

- It can (and often should) be performed
- Avoid testing hypothesis by looking at whether the credible interval crosses zero or not
- Better procedures exist:
 - Bayes Factors (Lee & Wagenmakers, 2013)
 - ROPE: Region of practical equivalence (Kruschke, 2014)

Bayes factors

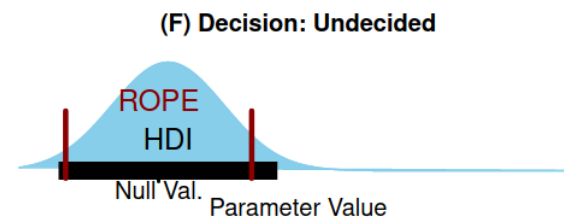
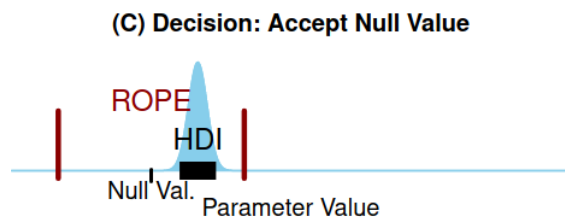
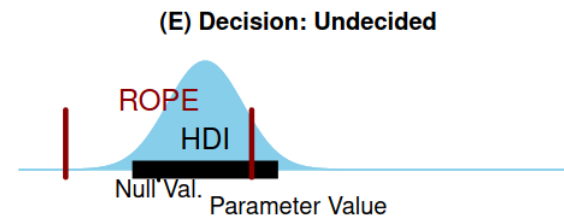
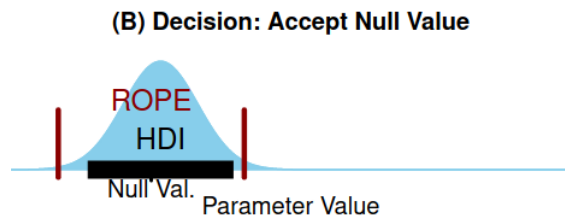
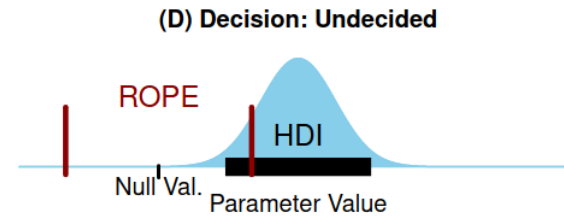
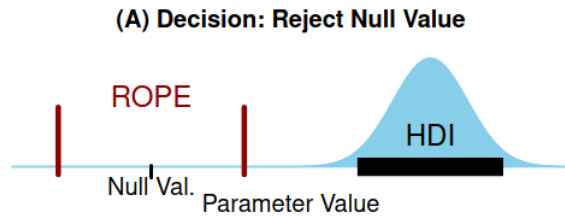
Lee and Wagenmakers (2013)



(figure by Kelter, 2020)

ROPE

Kruschke (2014)



Bayesian hypothesis testing

- Importantly, Bayesian hypothesis testing allows assessing the evidence both *for* and *against* a hypothesis
- Unlike with frequentist statistics, we can obtain support for *equality* between conditions or groups (non-significant p -values are inconclusive)

Bayesian statistics for bilingualism research

- Bilingualism research is characterized by:
 - Effects of different magnitudes in different groups
 - Prior information from previous studies, with either monolinguals or bilinguals
 - Heterogeneity between and within individuals (which can differ in different groups)
 - The use of many varied data types, tasks, and measures
 - Research questions about differences, as well as equality