# Introduction to Bayesian linear regression with brms

Stefano Coretta

18/01/2020

[mention installation]

# Random variables

- We have a question about the world, so we collect data (sample from a population).
  - y $= (y_1, y_2, y_3, y_4, ..., y_n)$
- We want to know how the data (the sample) was generated.
- In probability theory, data is generated by a random variable $Y$.

# Random variables

- $Y$ is uncertain.
  - We can describe $Y$ as a probability distribution, expressed by a set of parameters $\Theta = (\theta_1, ..., \theta_n)$.
- Probability distributions:
  - $Normal(\mu, \sigma)$,
  - $Binomial(n, p)$,
  - ...

# Random variables

$$vot_i \sim Normal(\mu, \sigma)$$

$$voiced_i \sim Bernoulli(p)$$

$$DoubleDative_i \sim Poisson(\lambda)$$

# Frequentist vs Bayesian view

- Parameters: $\mu$, $\sigma$, $p$, $\lambda$, ...
- Frequentist view:
  - The parameters are **fixed** (they are unknown but certain).
  - They take on a specific value.
- Bayesian view:
  - The parameters are **random variables** (they are unkown and uncertain).
  - We describe each parameter as a probability distribution, expressed by a set of **hyperparameters**.

# Continous random variable

$$vot_i \sim Normal(\mu, \sigma)$$

$$\mu \sim Normal(\mu_1, \sigma_1)$$

$$\sigma \sim HalfCauchy(x_0, \gamma)$$

# Bayes' Theorem

$$P(\theta \mid d) = \frac{P(d \mid \theta)\, P(\theta)}{P(d)}$$

# Bayes' Theorem

$$posterior\ probability = \frac{likelihood \times prior}{marginal\ likelihood}$$

# Priors

- We can incorporate previous knowledge about the hyperparameters as **priors** (prior distributions).
- Priors are chosen based on expert knowledge, previous studies, pilot data...
    - Priors must **not** be chosen based on the data to be analysed.

# Priors

- Informative and weakly informative priors.
- Uninformative or diffuse priors.
    - Uniform distribution.
- Regularising priors.

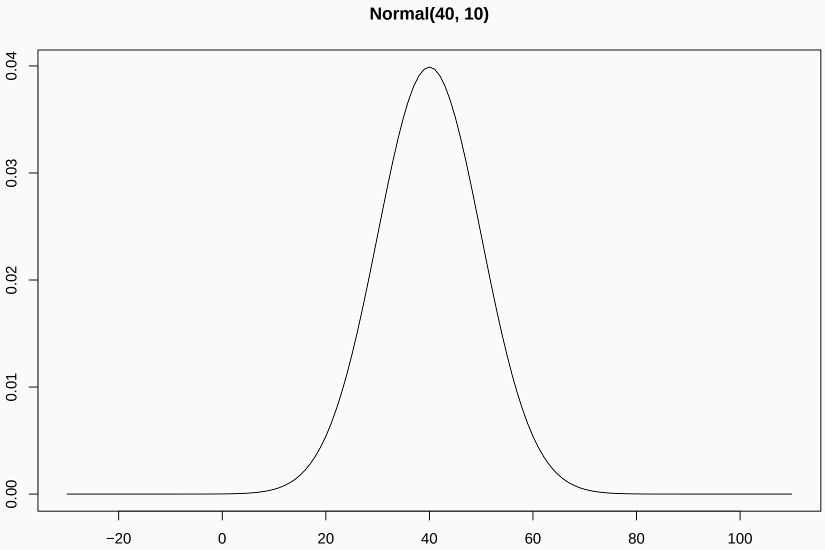
# Normal prior

[empirical rule]

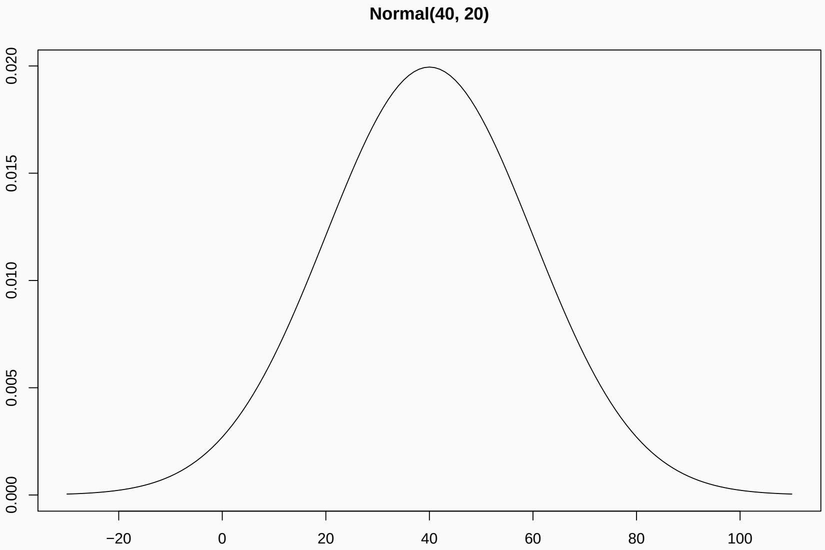
# Italian VOT

- Previous literature on VOT in Italian (Esposito, 2002; Stevens & Hajek, 2010) report VOT values for voiceless stops in the range of 20–60 ms.
    - We can express this knowledge with the prior $Normal(40, 10)$.
    - This is a somewhat strongly informative prior.

# Italian VOT



# Italian VOT



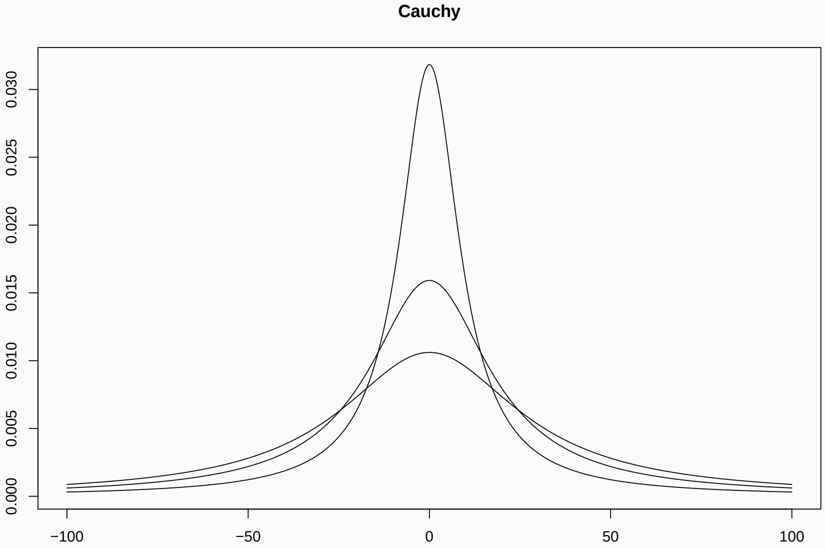
# Italian VOT

$$vot_i \sim Normal(\mu, \sigma)$$

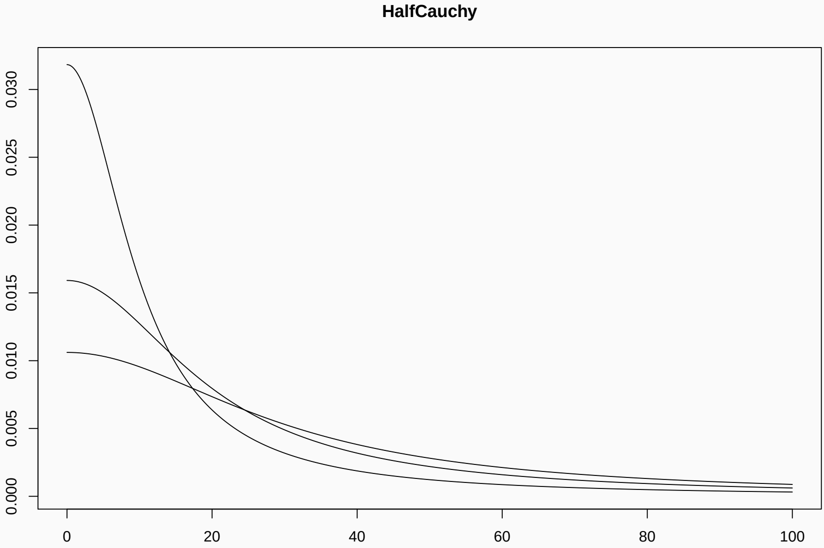
$$\mu \sim Normal(40, 10)$$

$$\sigma \sim HalfCauchy(x_0, \gamma)$$

# Cauchy prior



# Cauchy prior



# Italian VOT

$$vot_i \sim Normal(\mu, \sigma)$$

$$\mu \sim Normal(40, 10)$$

$$\sigma \sim HalfCauchy(0, 10)$$

# Italian VOT

- We have a model which incorporates (some of) our knowledge about VOT (through the priors for $\mu$ and $\sigma$).
- Now we want to obtain the **posterior distributions** of $\mu$ and $\sigma$.
    - The posterior distribution is the prior distribution *conditioned* on the data.
- **brms** R package: Bayesian Regression Models using Stan (Bürkner, 2018).

# brms

- Stan (Stan Development Team, 2017).
  - Statistical programming language written in C++ for fitting Bayesian models (calculate posterior distributions).
  - Calculation can be complex and/or impossible, so we take many samples from the data and from the possible parameter values to find the posterior distributions of the hyperparameters.
  - Markov Chain Monte Carlo (MCMC) sampling using the No-U-Turn sampler (NUTS).
- brms is an interface between R and Stan.
- `brm()` function from brms.
  - lme4 syntax (`y ~ x + (1|w)`).
  - Creates a Stan model, which is compiled and run.

# brms

```r
library(brms)

vot1 <- brm(
  <model_formula>,
  <family>,
  <prior>,
  <data>,
  chains = 4,
  iter = 2000
)
```

# brms

```r
library(brms)

vot1 <- brm(
  vot ~ 1,
  family = gaussian(),
  <prior>,
  data = ita_egg,
  chains = 4,
  iter = 2000
)
```

# Get prior

```r
get_prior(
  vot ~ 1,
  family = gaussian(),
  data = ita_egg
)
```

```
##                   prior     class coef group resp dpa
## 1 student_t(3, 19, 14) Intercept
## 2  student_t(3, 0, 14)     sigma
```

# Prior predictive checks

# Set prior

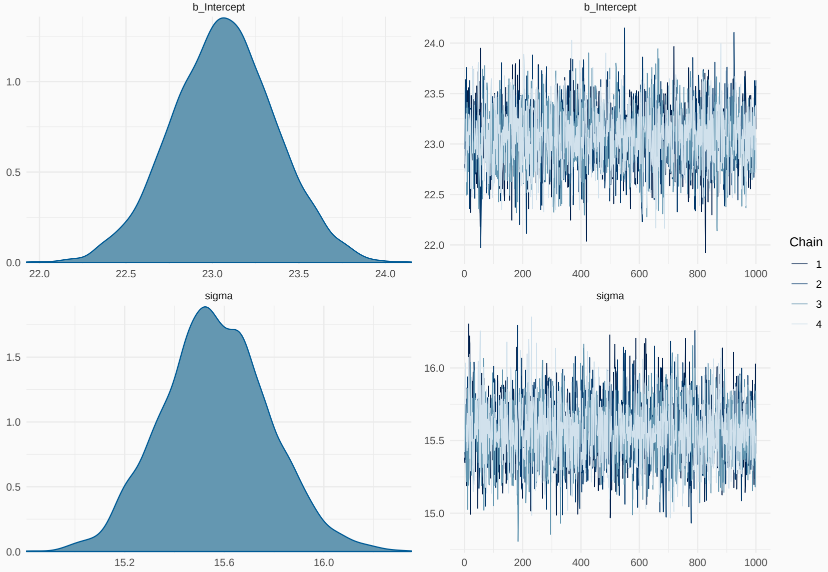
# Run the model

```
vot1 <- brm(
  vot ~ 1,
  family = gaussian(),
  prior = priors,
  data = ita_egg,
  chains = 4,
  iter = 2000,
  file = "./cache/vot1"
)
```

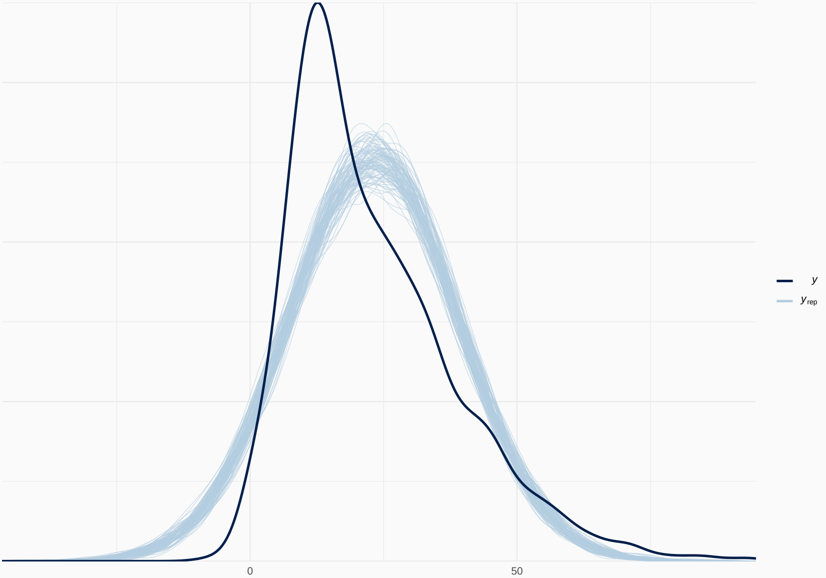
# Model summary

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: vot ~ 1
##    Data: ita_egg (Number of observations: 2624)
## Samples: 4 chains, each with iter = 2000; warmup = 
##          total post-warmup samples = 4000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat
## Intercept    23.06      0.30    22.47    23.65 1.00
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bul
## sigma    15.57      0.21    15.16    15.99 1.00
```

# Plot model


b_Intercept
1.0
0.5
0.0
22.0
22.5
23.0
23.5
24.0
b_Intercept
24.0
23.5
23.0
22.5
22.0
0
200
400
600
800
1000
Chain
1
2
3
4
sigma
1.5
1.0
0.5
0.0
15.2
15.6
16.0
sigma
16.0
15.5
15.0
0
200
400
600
800
1000

# Posterior predictive check



# References

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59(4). 197–231. doi:10.1159/000068347.

Stan Development Team. 2017. Stan: A C++ library for probability and sampling, version 2.14.0. http://mc-stan.org/.

Stevens, Mary & John Hajek. 2010. Post-aspiration in standard Italian: some first cross-regional acoustic evidence. Paper presented at Interspeech, 26-30 September 2010, Makuhari, Chiba, Japan.