Introduction to Bayesian linear regression with brms

Stefano Coretta

18/01/2020

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:
    - $Normal(\mu, \sigma)$,
    - $Binomial(n, p)$,
    - …

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

$$voiced \sim Bernoulli(p)$$

$$DoubleDative \sim Poisson(\lambda)$$

Frequentist vs Bayesian view

- Parameters: μ, σ, p, λ, …
- 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

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

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

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

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 uninformative priors.
- Informative priors.
    - Weakly informative.
    - Strongly informative.
- Uninformative priors.
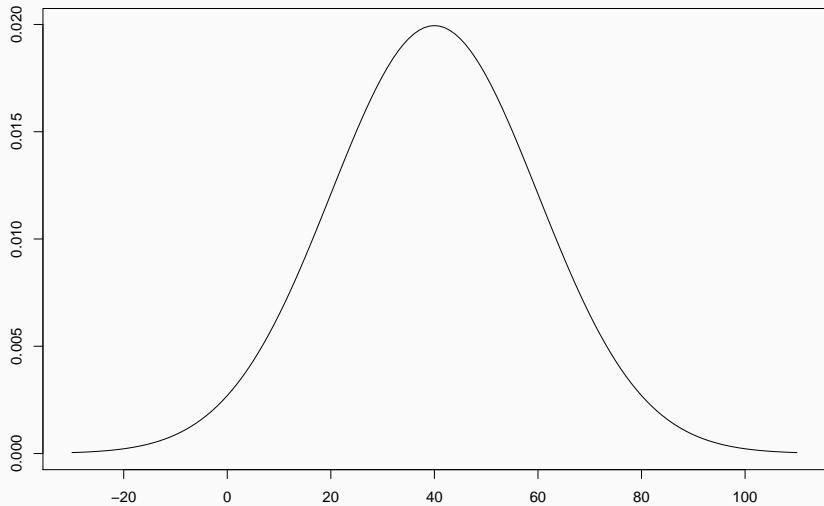   - Uniform distribution.

Normal prior

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
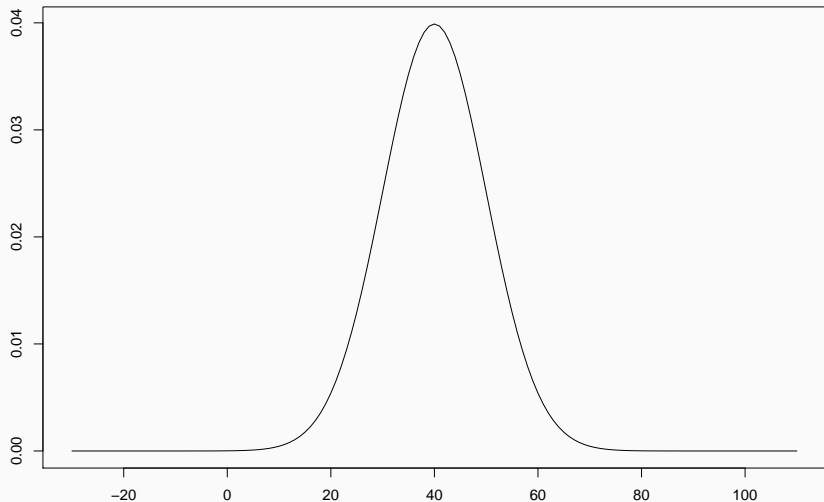
Italian VOT

References

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

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.