A few Bayesian lectures for the Uninitiated

Carsten F. Dormann

Biometry & Environmental System Analysis, University of Freiburg

21. Januar 2022

Why Bayes?

- likelihood $\mathcal{L}$ (and thus GLM, ANOVA and t-test) does not actually address the question we are interested in:
  $P($ hypothesis $\mid$ data $) \neq P($ data $\mid$ hypothesis $)$ $(= \mathcal{L})$
  Use Bayes Theorem to convert! [prosecutor's fallacy]

Why Bayes?

- likelihood $\mathcal{L}$ (and thus GLM, ANOVA and t-test) does not actually address the question we are interested in:
  $P(\text{ hypothesis } | \text{ data }) \neq P(\text{ data } | \text{ hypothesis }) \; (= \mathcal{L})$
  Use Bayes Theorem to convert! [prosecutor's fallacy]
- Research is rarely unprecedented: We should use previous knowledge! [mind reading, tachyons]

Why Bayes?

- likelihood $\mathcal{L}$ (and thus GLM, ANOVA and t-test) does not actually address the question we are interested in:
  $P(\text{ hypothesis } | \text{ data }) \neq P(\text{ data } | \text{ hypothesis }) \, (= \mathcal{L})$
  Use Bayes Theorem to convert! [prosecutor's fallacy]
- Research is rarely unprecedented: We should use previous knowledge! [mind reading, tachyons]
- Things are not always linear: GLMs are limited. Things are not always simple: GLMs are. [non-linear; POM]

Why Bayes?

- likelihood $\mathcal{L}$ (and thus GLM, ANOVA and t-test) does not actually address the question we are interested in: $P(\text{ hypothesis} \mid \text{data }) \neq P(\text{ data} \mid \text{hypothesis })\ (= \mathcal{L})$ Use Bayes Theorem to convert! [prosecutor's fallacy]
- Research is rarely unprecedented: We should use previous knowledge! [mind reading, tachyons]
- Things are not always linear: GLMs are limited. Things are not always simple: GLMs are. [non-linear; POM]
- `lme4::glmer` and friends keep throwing error messages: fragile estimation procedure; Bayesian algorithms more robust [MCMC as optimiser]

Why Bayes?

- likelihood $\mathcal{L}$ (and thus GLM, ANOVA and t-test) does not actually address the question we are interested in:
  $P(\text{ hypothesis} \mid \text{data }) \neq P(\text{ data} \mid \text{hypothesis }) \; (= \mathcal{L})$
  Use Bayes Theorem to convert! [prosecutor's fallacy]
- Research is rarely unprecedented: We should use previous knowledge! [mind reading, tachyons]
- Things are not always linear: GLMs are limited. Things are not always simple: GLMs are. [non-linear; POM]
- `lme4::glmer` and friends keep throwing error messages: fragile estimation procedure; Bayesian algorithms more robust [MCMC as optimiser]
- (Fine-tune shrinking against collinearity, for model selection.)
- (Get interpretable "credible intervals".)

Why **not** Bayes? Or: why your colleagues will hate you.

- Another new statistical machismo to disenfranchise non-geeks.

Why **not** Bayes? Or: why your colleagues will hate you.

- Another new statistical machismo to disenfranchise non-geeks.
- Editors don't like it in non-Bayes journals (because they are not familiar enough with it).

Why **not** Bayes? Or: why your colleagues will hate you.

- Another new statistical machismo to disenfranchise non-geeks.
- Editors don't like it in non-Bayes journals (because they are not familiar enough with it).
- Coauthors don't like the tedious and long method section and cannot readily re-run the analysis (because they are not familiar enough with it).

Why **not** Bayes? Or: why your colleagues will hate you.

- Another new statistical machismo to disenfranchise non-geeks.
- Editors don't like it in non-Bayes journals (because they are not familiar enough with it).
- Coauthors don't like the tedious and long method section and cannot readily re-run the analysis (because they are not familiar enough with it).
- Priors require change in research mentality, so currently it looks like fudging your analysis into what you want to find (while in fact it is exactly the opposite).

Why **not** Bayes? Or: why your colleagues will hate you.

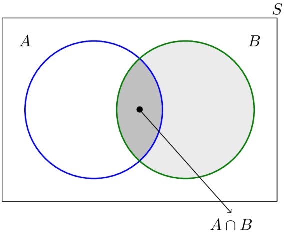
- Another new statistical machismo to disenfranchise non-geeks.
- Editors don't like it in non-Bayes journals (because they are not familiar enough with it).
- Coauthors don't like the tedious and long method section and cannot readily re-run the analysis (because they are not familiar enough with it).
- Priors require change in research mentality, so currently it looks like fudging your analysis into what you want to find (while in fact it is exactly the opposite).
- Everything takes ages and we make a lot of mistakes (because we are not familiar enough with it).

Why **not** Bayes? Or: why your colleagues will hate you.

- Another new statistical machismo to disenfranchise non-geeks.
- Editors don't like it in non-Bayes journals (because they are not familiar enough with it).
- Coauthors don't like the tedious and long method section and cannot readily re-run the analysis (because they are not familiar enough with it).
- Priors require change in research mentality, so currently it looks like fudging your analysis into what you want to find (while in fact it is exactly the opposite).
- Everything takes ages and we make a lot of mistakes (because we are not familiar enough with it).

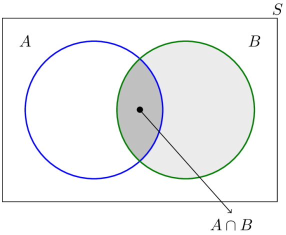
These *cannot* be valid arguments, otherwise we'll never see progress in science!

Deriving Bayes Theorem



S is a buffet, with warm dishes (A) and desserts (B). What is you chance of randomly grabbing a warm dessert?

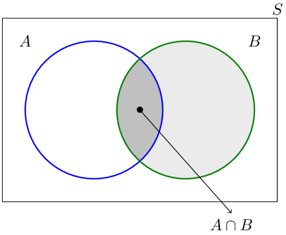
Deriving Bayes Theorem



S is a buffet, with warm dishes (A) and desserts (B). What is you chance of randomly grabbing a warm dessert?

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B)P(A|B)$$

Deriving Bayes Theorem



S is a buffet, with warm dishes (A) and desserts (B). What is you chance of randomly grabbing a warm dessert?

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B)P(A|B)$$

Re-arrange:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \tag{1}$$

Final slide: Prosecutor's Fallacy

In court, a forensic scientist states that the probability that an innocent has a DNA-match to the sample found at a crime scene is 1:3 million:

$$P(\text{DNA-match}|\text{innocent}) = 1/3\text{mio}.$$

The accused has a DNA-match: a clear case! – Or is it?

Final slide: Prosecutor's Fallacy

In court, a forensic scientist states that the probability that an innocent has a DNA-match to the sample found at a crime scene is 1:3 million:

$$P(\text{DNA-match}|\text{innocent}) = 1/3\text{mio}.$$

The accused has a DNA-match: a clear case! – Or is it?

No! What we actually want to know: Probability of being innocent when the DNA matches:

$$P(\text{innocent}|\text{DNA-match})$$

Final slide: Prosecutor's Fallacy

In court, a forensic scientist states that the probability that an innocent has a DNA-match to the sample found at a crime scene is 1:3 million:

$$P(\text{DNA-match}|\text{innocent}) = 1/3\text{mio}.$$

The accused has a DNA-match: a clear case! – Or is it?

No! What we actually want to know: Probability of being innocent when the DNA matches:

$$P(\text{innocent}|\text{DNA-match}) \neq P(\text{DNA-match}|\text{innocent})$$

Final slide: Prosecutor's Fallacy

In court, a forensic scientist states that the probability that an innocent has a DNA-match to the sample found at a crime scene is 1:3 million:

$$P(\text{DNA-match}|\text{innocent}) = 1/3\text{mio}.$$

The accused has a DNA-match: a clear case! – Or is it?

No! What we actually want to know: Probability of being innocent when the DNA matches:

$$P(\text{innocent}|\text{DNA-match}) \neq P(\text{DNA-match}|\text{innocent})$$

Bring in the Bayes Theorem!

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \tag{2}$$

[I lied:] Final final slide: Prosecutor's Fallacy

$$P(\text{innocent}|\text{DNA-match}) = \frac{P(\text{DNA-match}|\text{innocent})P(\text{innocent})}{P(\text{DNA-match})} \tag{3}$$

If there are 10 million people in the DNA database, we'll get:

$$P(\text{innocent}|\text{DNA-match}) = \frac{(1/3 \cdot 10^6) \cdot (10^7 - 1)}{(1/3 \cdot 10^6 \cdot 10^7) + 1}$$

[I lied:] Final final slide: Prosecutor's Fallacy

$$P(\text{innocent}|\text{DNA-match}) = \frac{P(\text{DNA-match}|\text{innocent})P(\text{innocent})}{P(\text{DNA-match})} \tag{3}$$

If there are 10 million people in the DNA database, we'll get:

$$\begin{aligned} P(\text{innocent}|\text{DNA-match}) &= \frac{(1/3 \cdot 10^6) \cdot (10^7 - 1)}{(1/3 \cdot 10^6 \cdot 10^7) + 1} \\ &= \frac{3.33}{4.33} = 0.77 \end{aligned}$$

[I lied:] Final final slide: Prosecutor's Fallacy

$$P(\text{innocent}|\text{DNA-match}) = \frac{P(\text{DNA-match}|\text{innocent})P(\text{innocent})}{P(\text{DNA-match})} \tag{3}$$

If there are 10 million people in the DNA database, we'll get:

$$\begin{aligned} P(\text{innocent}|\text{DNA-match}) &= \frac{(1/3 \cdot 10^6) \cdot (10^7 - 1)}{(1/3 \cdot 10^6 \cdot 10^7) + 1} \\ &= \frac{3.33}{4.33} = 0.77 \end{aligned}$$

Bayes Theorem shows us how to do the maths.

Next time:
Turning a little GLM-style regression into something weirdly complicated, just to call it Bayesian.