Basics

Applied Bayesian Statistics
Winter Term 2018

Susumu Shikano
GSDS

This session

Goals

- Basic logic and concepts of Bayesian inference.
- Likelihood, prior and posterior
- Conjugate analysis

An example research question

Which ideological orientation does this student candidate have? Left or right?

An example research question

Which ideological orientation does this student candidate have? Left or right?

An example research question

Which ideological orientation does this student candidate have? Left or right?

An example research question

Which ideological orientation does this student candidate have? Left or right?

An example research question

Which ideological orientation does this student candidate have? Left or right?

- Right (RCDS)
- Left (JUSO)

An example research question

Now assume. . .

- A person claims that he should be able to distinguish ideological orientation of others (left or right) only based on the facial characteristics in 7 out of 10 cases.

An example research question

Now assume. . .

- A person claims that he should be able to distinguish ideological orientation of others (left or right) only based on the facial characteristics in 7 out of 10 cases.
- You are sceptical.

An example research question

Now assume...

- A person claims that he should be able to distinguish ideological orientation of others (left or right) only based on the facial characteristics in 7 out of 10 cases.
- You are sceptical.
    - If you are frequentist: Let him judge the ideological orientation of many people.
    - Given: He could correctly judge the ideological orientation of 60 out of 100 persons.

If you are frequentist. . .

you take a model with the maximum likelihood.

- There are two models
  - H_1: He can distinguish 7 out of 10.
  - H_0: His answer is random.

If you are frequentist...

you take a model with the maximum likelihood.

- There are two models
  - H_1: He can distinguish 7 out of 10.
  - H_0: His answer is random.
- ...and you calculate how likely the data (y) is collected under different models.
- $\rightarrow$ Likelihood

If you are frequentist...

Computing Likelihood ($p(y|\theta)$)

If you are frequentist...

Computing Likelihood ($p(y|\theta)$)

- Which probability distribution form has a percentage?

If you are frequentist...

Computing Likelihood ($p(y|\theta)$)

- Which probability distribution form has a percentage?
- Binomial distribution

If you are frequentist...

Computing Likelihood ($p(y|\theta)$)

- Which probability distribution form has a percentage?
- Binomial distribution

Bernoulli distribution

Coin toss: If the coin is fair,

$$\begin{aligned} \Pr(Y_i = \text{``}head''\text{)} &= 0.5 \\ \Pr(Y_i = \text{``}tail''\text{)} &= 0.5 \\ \Pr(Y_i \neq \text{``}head'', \text{``}tail''\text{)} &= 0 \end{aligned}$$

If you are frequentist...

Bernoulli distribution

More generally

$$
\begin{aligned}
\Pr(Y_i = 1) &= \pi \\
\Pr(Y_i = 0) &= 1 - \pi \\
\Pr(Y_i \neq 0, 1) &= 0
\end{aligned}
$$

- parameter: π
- $\pi \in [0, 1]$

This is called Bernoulli distribution.

Formally

$$
Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}
$$

If you are frequentist...

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

If you are frequentist. . .

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

How likely do you have heads

- for three times?: $\Pr(Y_i = 3) = \pi^3$

If you are frequentist...

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

How likely do you have heads

- for three times?: $\Pr(Y_i = 3) = \pi^3$
- never?: $\Pr(Y_i = 0) = (1-\pi)^3$

If you are frequentist...

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

How likely do you have heads

- for three times?: $\Pr(Y_i = 3) = \pi^3$
- never?: $\Pr(Y_i = 0) = (1-\pi)^3$
- for twice?: $\Pr(Y_i = 2) = \pi^2(1-\pi)$?

If you are frequentist...

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

How likely do you have heads

- for three times?: $\Pr(Y_i = 3) = \pi^3$
- never?: $\Pr(Y_i = 0) = (1-\pi)^3$
- for twice?: $\Pr(Y_i = 2) = \pi^2(1-\pi)$?
- ...No! $\Pr(Y_i = 2) = 3\pi^2(1-\pi)$

If you are frequentist...

Binomial distribution

You repeat coin tossing for three times (N=3).
Important assumption: each trial is independent and identical.

$$Y_i \sim f_{Bern}(y_i|\pi) = \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i} & \text{for } y_i = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

How likely do you have heads

- for three times?: $\Pr(Y_i = 3) = \pi^3$
- never?: $\Pr(Y_i = 0) = (1-\pi)^3$
- for twice?: $\Pr(Y_i = 2) = \pi^2(1-\pi)$?
- ...No! $\Pr(Y_i = 2) = 3\pi^2(1-\pi)$

$$3 = \frac{3 \times 2}{2 \times 1} = \begin{pmatrix} N \\ y_i \end{pmatrix} = \frac{N!}{y_i!(N-y_i)!}$$

If you are frequentist...

Binomial distribution

Generally...

$$
\begin{aligned}
f_b(y_i|\pi, N) &= \Pr(Y_i = y_i|\pi, N) \\
&= \begin{cases} \begin{pmatrix} N \\ y_i \end{pmatrix} \pi^{y_i}(1-\pi)^{N-y_i} & \text{for } y_i = 0, 1, \ldots, N \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

If you are frequentist...

Back to our example

- There are two models
  - H_1: He can distinguish 7 out of 10. $\rightarrow \pi = 0.7$
  - H_0: His answer is random. $\rightarrow \pi = 0.5$

If you are frequentist...

Back to our example

- There are two models
  - H_1: He can distinguish 7 out of 10. $\rightarrow \pi = 0.7$
  - H_0: His answer is random. $\rightarrow \pi = 0.5$
- ...and you calculate how likely the data (y) is collected under different models.

If you are frequentist...

Back to our example

- There are two models
  - H_1: He can distinguish 7 out of 10. $\rightarrow \pi = 0.7$
  - H_0: His answer is random. $\rightarrow \pi = 0.5$
- ...and you calculate how likely the data (y) is collected under different models.

$$\begin{aligned}\Pr(y|H_1) &= \binom{100}{60} \cdot 0.7^{60} \cdot 0.3^{40} \\ &\approx 0.0085\end{aligned}$$

If you are frequentist...

Back to our example

- There are two models
  - H_1: He can distinguish 7 out of 10. $\rightarrow \pi = 0.7$
  - H_0: His answer is random. $\rightarrow \pi = 0.5$
- ...and you calculate how likely the data (y) is collected under different models.

$$
\begin{aligned}
\Pr(y|H_1) &= \begin{pmatrix} 100 \\ 60 \end{pmatrix} \cdot 0.7^{60} \cdot 0.3^{40} \\
&\approx 0.0085
\end{aligned}
$$

$$
\begin{aligned}
\Pr(y|H_0) &= \begin{pmatrix} 100 \\ 60 \end{pmatrix} \cdot 0.5^{60} \cdot 0.5^{40} \\
&\approx 0.0108 \rightarrow \text{larger likelihood!}
\end{aligned}
$$

$* \; y = \{60 \text{ times right; } 40 \text{ times wrong}\}$

If you are Bayesian...

What is at stake?: Your belief about how likely the models are!!

- $\Pr(H)$ instead of $\Pr(y)$.

If you are Bayesian...

What is at stake?: Your belief about how likely the models are!!

- $\Pr(H)$ instead of $\Pr(y)$.
- Bayes Theorem!

$$\underbrace{p(H|y)}_{\textit{posterior}} = \frac{\overbrace{p(y|H)}^{\textit{likelihood}}\, p(H)}{p(y)}$$

If you are Bayesian. . .

What is at stake?: Your belief about how likely the models are!!

- $\Pr(H)$ instead of $\Pr(y)$.
- Bayes Theorem!

$$\underbrace{p(H|y)}_{\textit{posterior}} = \frac{\overbrace{p(y|H)}^{\textit{likelihood}}\, p(H)}{p(y)}$$

- $\Pr(H)$??

If you are Bayesian…

What is at stake?: Your belief about how likely the models are!!

- $\Pr(H)$ instead of $\Pr(y)$.
- Bayes Theorem!

$$\underbrace{p(H|y)}_{\textit{posterior}} = \frac{\overbrace{p(y|H)}^{\textit{likelihood}}\, p(H)}{p(y)}$$

- $\Pr(H)$??
- The belief which your have about models prior to the data collection.

If you are Bayesian...

What is at stake?: Your belief about how likely the models are!!

- $\Pr(H)$ instead of $\Pr(y)$.
- Bayes Theorem!

$$\underbrace{p(H|y)}_{\textit{posterior}} = \frac{\overbrace{p(y|H)}^{\textit{likelihood}}\, p(H)}{p(y)}$$

- $\Pr(H)$??
- The belief which your have about models prior to the data collection.
- E.g.: Before testing, it is 50:50 whether he or you are right. There can be no further model ($\Pr(H_1) + \Pr(H_0) = 1$).
- $\rightarrow \Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$

If you are Bayesian...

$\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$

- The first trial: successful judgement.

If you are Bayesian...

$\Pr(H_1) = 0.5$ **and** $\Pr(H_0) = 0.5$

- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)}
\end{aligned}
$$

If you are Bayesian...

$\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$

- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)}
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)}
$$

If you are Bayesian...

$\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$

- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)} = \frac{0.35}{0.35 + 0.25} \approx 0.583
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)} = \frac{0.25}{0.35 + 0.25} \approx 0.417
$$

If you are Bayesian...

$\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$

- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)} = \frac{0.35}{0.35 + 0.25} \approx 0.583
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)} = \frac{0.25}{0.35 + 0.25} \approx 0.417
$$

- Even after only one observation you can say something about the models.

If you are Bayesian...

After first trial: $\Pr(H_1) = 0.583$ **and** $\Pr(H_0) = 0.417$

- The next trial: successful judgement again.

$$p(H_1|y) = \frac{p(y|H_1)p(H_1)}{p(y)} = \frac{0.7^1 \cdot 0.3^0 \times 0.583}{p(y)}$$

$$p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.417}{p(y)}$$

If you are Bayesian...

After first trial: $\Pr(H_1) = 0.583$ **and** $\Pr(H_0) = 0.417$

- The next trial: successful judgement again.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{0.7^1 \cdot 0.3^0 \times 0.583}{p(y)} = \frac{0.408}{p(y)} = \frac{0.408}{0.408 + 0.209} \approx 0.661
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.417}{p(y)} = \frac{0.209}{p(y)} = \frac{0.209}{0.408 + 0.209} \approx 0.339
$$

If you are Bayesian...

After first trial: $\Pr(H_1) = 0.583$ **and** $\Pr(H_0) = 0.417$

- The next trial: successful judgement again.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{0.7^1 \cdot 0.3^0 \times 0.583}{p(y)} = \frac{0.408}{p(y)} = \frac{0.408}{0.408 + 0.209} \approx 0.661
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.417}{p(y)} = \frac{0.209}{p(y)} = \frac{0.209}{0.408 + 0.209} \approx 0.339
$$

- The posterior is different after the same observation (successful judgement).
- You are more sure that he is right (H_1 is correct).

If you are Bayesian. . .

After 100 trials with 60 correct judgements. . .

$$p(H_1|y) = \frac{p(y|H_1)p(H_1)}{p(y)}$$

$$= \frac{0.0085 \times 0.5}{p(y)} = \frac{0.0425}{p(y)} = \frac{0.0425}{0.0425 + 0.0504} \approx 0.439$$

$$p(H_0|y) = \frac{0.0108 \times 0.5}{p(y)} = \frac{0.0504}{p(y)} = \frac{0.0504}{0.0425 + 0.0504} \approx 0.561$$

If you are Bayesian...

After 100 trials with 60 correct judgements...

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{0.0085 \times 0.5}{p(y)} = \frac{0.0425}{p(y)} = \frac{0.0425}{0.0425 + 0.0504} \approx 0.439
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.0108 \times 0.5}{p(y)} = \frac{0.0504}{p(y)} = \frac{0.0504}{0.0425 + 0.0504} \approx 0.561
$$

- You tend to believe more H_0 than H_1.

If you are Bayesian...

After 100 trials with 60 correct judgements...

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{0.0085 \times 0.5}{p(y)} = \frac{0.0425}{p(y)} = \frac{0.0425}{0.0425 + 0.0504} \approx 0.439 \\
p(H_0|y) &= \frac{0.0108 \times 0.5}{p(y)} = \frac{0.0504}{p(y)} = \frac{0.0504}{0.0425 + 0.0504} \approx 0.561
\end{aligned}
$$

- You tend to believe more H_0 than H_1.
- What if we have further models besides H_0 and H_1?
  - He can distinguish 6 out of 10, 8 out of 10,...

If you are Bayesian...

After 100 trials with 60 correct judgements...

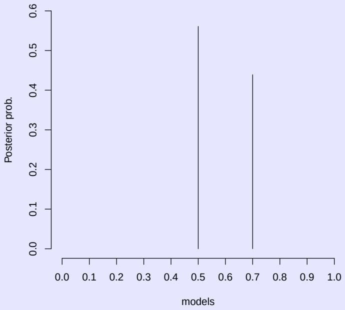
$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{0.0085 \times 0.5}{p(y)} = \frac{0.0425}{p(y)} = \frac{0.0425}{0.0425 + 0.0504} \approx 0.439
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.0108 \times 0.5}{p(y)} = \frac{0.0504}{p(y)} = \frac{0.0504}{0.0425 + 0.0504} \approx 0.561
$$

- You tend to believe more H_0 than H_1.
- What if we have further models besides H_0 and H_1?
  - He can distinguish 6 out of 10, 8 out of 10,...
- You can extend the calculation so many times as you like.

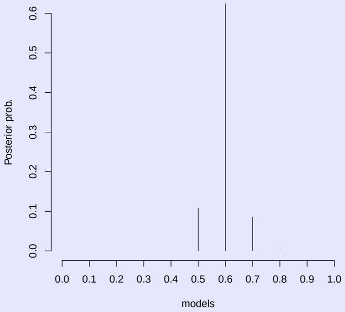
If you are Bayesian...

Not only two models, but more models



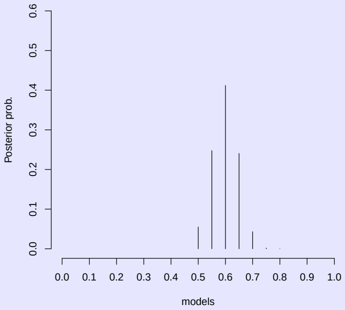
If you are Bayesian. . .

Not only two models, but more models

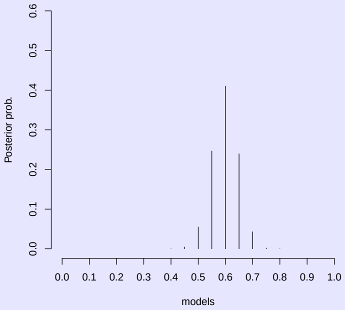


If you are Bayesian...

Not only two models, but more models

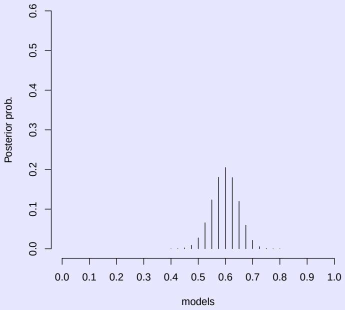


Not only two models, but more models



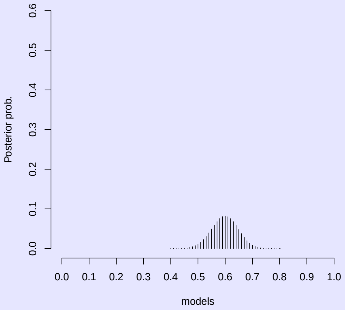
If you are Bayesian. . .

Not only two models, but more models



If you are Bayesian...

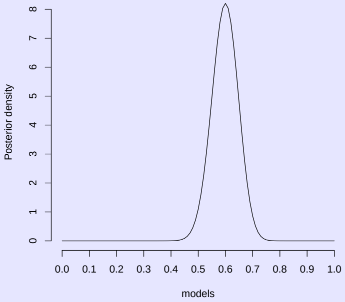
Not only two models, but more models



- However, it is often the case that we have uncountable number of models (or better: parameter values)...

If you are Bayesian...

Not only two models, but more models

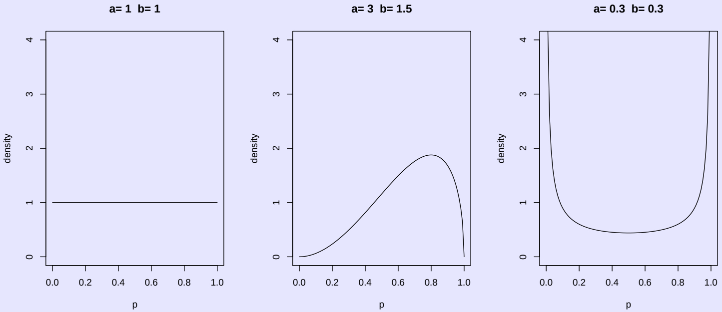


- However, it is often the case that we have uncountable number of models (or better: parameter values)...
- We can solve the problem analytically (conjugate analysis).

Conjugate analysis

One can use beta distribution as continuous prior...

Beta distribution: $f_\beta(\pi|a, b)$

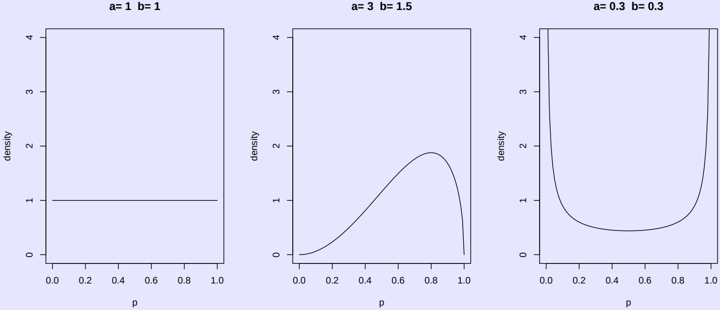


Why?

Conjugate analysis

One can use beta distribution as continuous prior...

Beta distribution: $f_\beta(\pi|a, b)$



Why?

- The relevant range (between 0 and 1)
- Flexible shapes

Conjugate analysis

Beta distribution: $f_\beta(\pi|a,b)$

$$\begin{aligned}
f_\beta(\pi|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \\
& \quad 0 \le \pi \le 1, a, b > 0 \\
E(\pi) &= \frac{a}{a+b} \\
Var(\pi) &= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

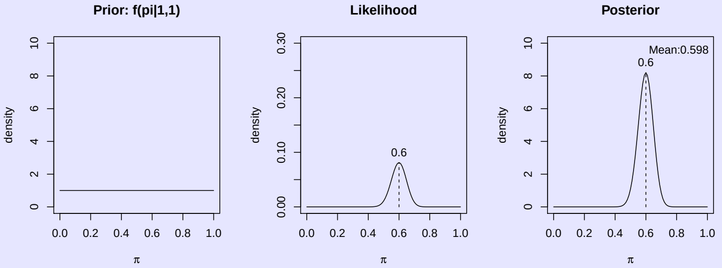
$*$ If n is a positive integer:

$$\Gamma(n) = (n-1)!$$

Conjugate analysis

If one has a continuous prior based on beta distribution...

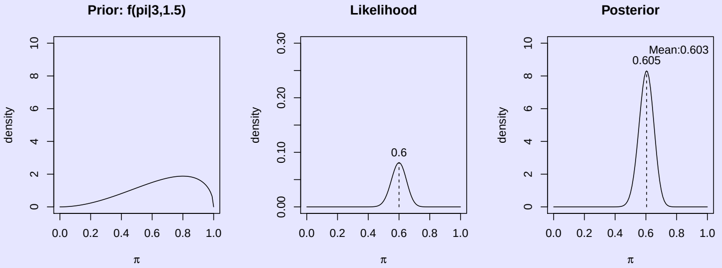
- Using beta priors
- and data (60 out of 100 are correct judgements),
- posterior can be analytically calculated.



Conjugate analysis

If one has a continuous prior based on beta distribution...

- Using beta priors
- and data (60 out of 100 are correct judgements),
- posterior can be analytically calculated.



Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f_\beta(\pi|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \\
f_b(y_i|\pi,N) &= \Pr(Y_i = y_i|\pi,N) \\
&= \begin{cases} \begin{pmatrix} N \\ y_i \end{pmatrix} \pi^{y_i}(1-\pi)^{N-y_i} & \text{for } y_i = 0,1,\ldots,N \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f_\beta(\pi|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \\
f_b(y_i|\pi,N) &= \Pr(Y_i = y_i|\pi,N) \\
&= \begin{cases} \begin{pmatrix} N \\ y_i \end{pmatrix} \pi^{y_i}(1-\pi)^{N-y_i} & \text{for } y_i = 0,1,\ldots,N \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- Both functions have a similar component.

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f_\beta(\pi|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \\
f_b(y_i|\pi,N) &= \Pr(Y_i = y_i|\pi,N) \\
&= \begin{cases} \begin{pmatrix} N \\ y_i \end{pmatrix} \pi^{y_i}(1-\pi)^{N-y_i} & \text{for } y_i = 0,1,\ldots,N \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- Both functions have a similar component.
- Beta distribution has normalizing constant.

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$f(\pi|y_i) = \frac{f_b(y_i|\pi) \cdot f_\beta(\pi)}{f(y_i)}$$

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f(\pi|y_i) &= \frac{f_b(y_i|\pi)\cdot f_\beta(\pi)}{f(y_i)} \\
&= \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}\cdot\begin{pmatrix} N \\ y_i \end{pmatrix}\pi^{y_i}(1-\pi)^{N-y_i}}{f(y_i)}
\end{aligned}
$$

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f(\pi|y_i) &= \frac{f_b(y_i|\pi)\cdot f_\beta(\pi)}{f(y_i)} \\
&= \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}\cdot\begin{pmatrix} N \\ y_i \end{pmatrix}\pi^{y_i}(1-\pi)^{N-y_i}}{f(y_i)} \\
&= \mathcal{C}\cdot\pi^{a+y_i-1}(1-\pi)^{b+N-y_i-1}
\end{aligned}
$$

Conjugate analysis

Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f(\pi|y_i) &= \frac{f_b(y_i|\pi)\cdot f_\beta(\pi)}{f(y_i)} \\
&= \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}\cdot\begin{pmatrix} N \\ y_i \end{pmatrix}\pi^{y_i}(1-\pi)^{N-y_i}}{f(y_i)} \\
&= \mathcal{C}\cdot\pi^{a+y_i-1}(1-\pi)^{b+N-y_i-1} \\
&= \frac{\Gamma(a+b+N)}{\Gamma(a+y_i)\Gamma(b+N-y_i)}\pi^{a+y_i-1}(1-\pi)^{b+N-y_i-1}
\end{aligned}
$$

Conjugate analysis

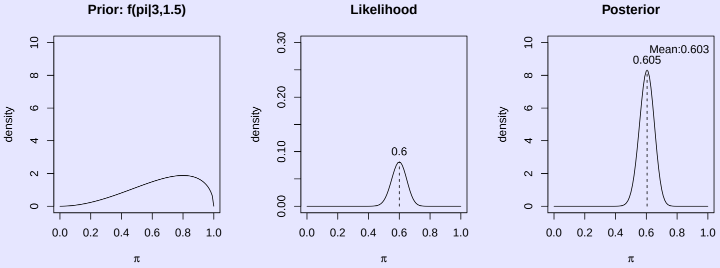
Calculating posterior by multiplying beta distribution and binomial distribution?

$$
\begin{aligned}
f(\pi|y_i) &= \frac{f_b(y_i|\pi) \cdot f_\beta(\pi)}{f(y_i)} \\
&= \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \cdot \begin{pmatrix} N \\ y_i \end{pmatrix} \pi^{y_i}(1-\pi)^{N-y_i}}{f(y_i)} \\
&= \mathcal{C} \cdot \pi^{a+y_i-1}(1-\pi)^{b+N-y_i-1} \\
&= \frac{\Gamma(a+b+N)}{\Gamma(a+y_i)\Gamma(b+N-y_i)}\pi^{a+y_i-1}(1-\pi)^{b+N-y_i-1} \\
&= f_\beta(\pi|a+y_i, b+N-y_i)
\end{aligned}
$$

- Prior and posterior have a same distribution form $\rightarrow$ Conjugacy

Conjugate analysis

If one has a continuous prior based on beta distribution...

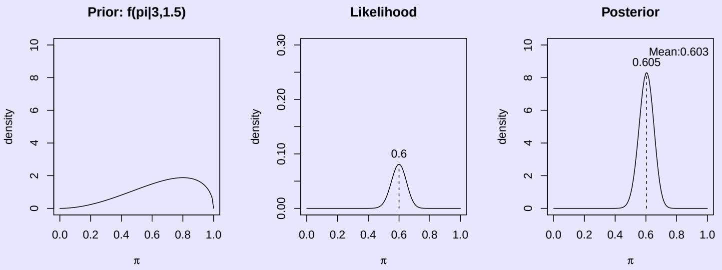


Or...

- prior: $p(\pi) = f_\beta(a, b)$

Conjugate analysis

If one has a continuous prior based on beta distribution...

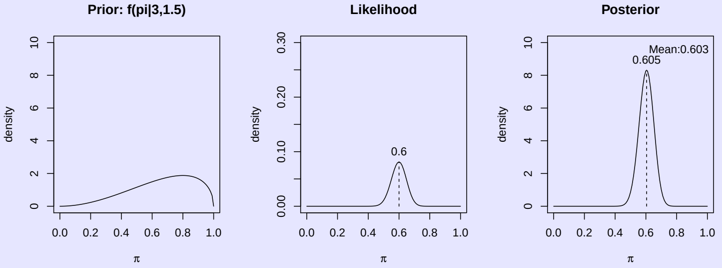


Or...

- prior: $p(\pi) = f_\beta(a, b)$
- likelihood: $f_b(60|\pi, 100)$

Conjugate analysis

If one has a continuous prior based on beta distribution...



Or...

- prior: $p(\pi) = f_\beta(a, b)$
- likelihood: $f_b(60|\pi, 100)$
- posterior: $p(\pi|y) = f_\beta(a + 60, b + 100 - 60)$

Conjugate analysis

Conjugacy

- Prior and posterior have a same distribution form.
- Easy to compute.

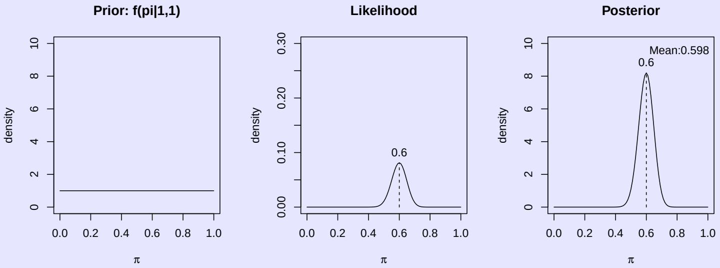
Conjugate analysis

Conjugacy

- Prior and posterior have a same distribution form.
- Easy to compute.
- Use of posterior as prior in subsequent analysis

Conjugate analysis

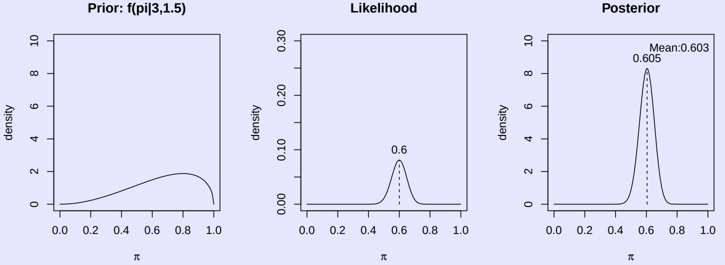
What we can observe here:



- The more certain $p(\theta)$ is,

Conjugate analysis

What we can observe here:

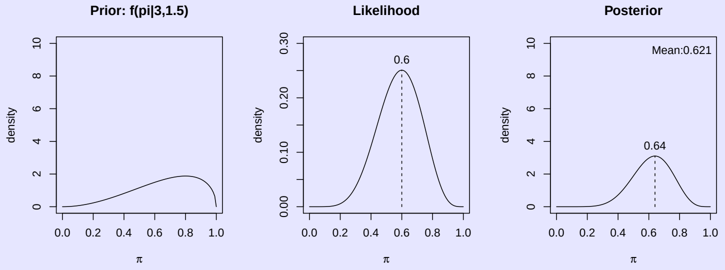


- The more certain $p(\theta)$ is,

...the more the prior distribution influences the posterior.
$\rightarrow$ A more different result from ML.

Conjugate analysis

What we can observe here:



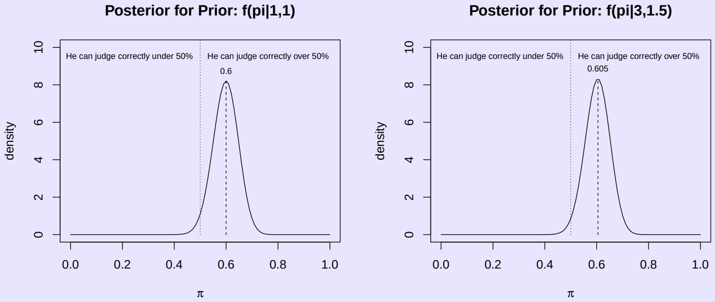
- The more certain $p(\theta)$ is,
- The smaller n is,*

…the more the prior distribution influences the posterior.
$\rightarrow$ A more different result from ML.

* Here, 6 out of 10 are correct judgement.

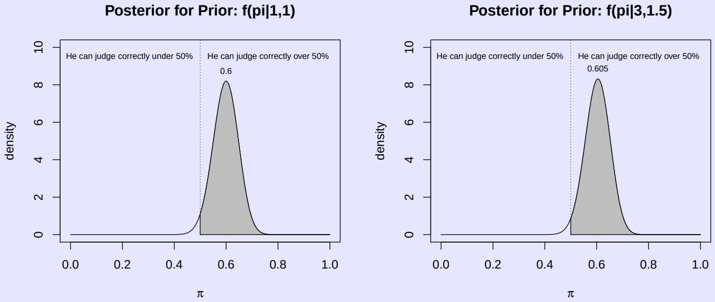
Conjugate analysis

What kind of statement?



Conjugate analysis

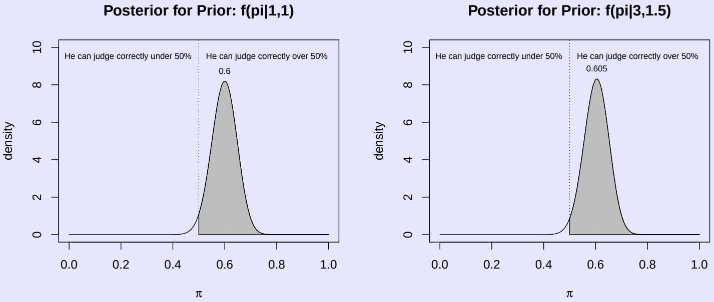
What kind of statement?



- You can report the posterior percentage of your model(s).

Conjugate analysis

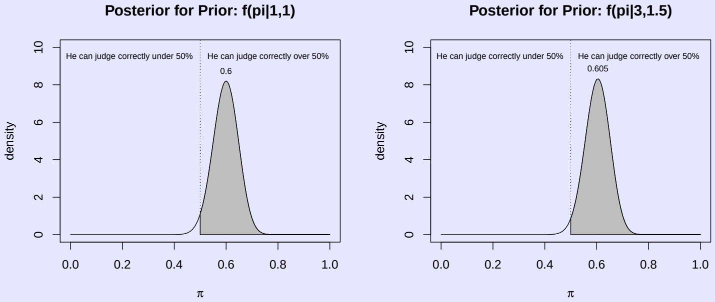
What kind of statement?



- You can report the posterior percentage of your model(s).
- But how can we calculate the percentage?

Conjugate analysis

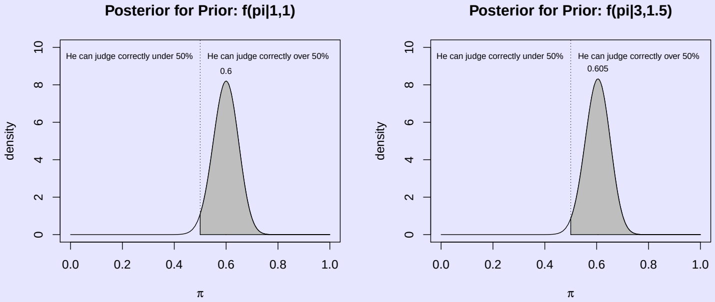
What kind of statement?



- You can report the posterior percentage of your model(s).
- But how can we calculate the percentage?
- You have to integrate the posterior.

Conjugate analysis

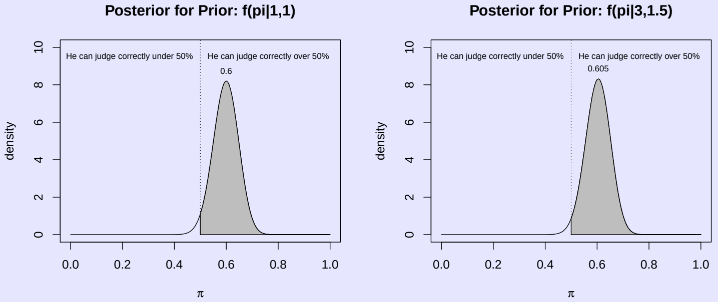
What kind of statement?



- You can report the posterior percentage of your model(s).
- But how can we calculate the percentage?
- You have to integrate the posterior.
- Sometimes difficult or unfeasible. . .

Conjugate analysis

What kind of statement?



- You can report the posterior percentage of your model(s).
- But how can we calculate the percentage?
- You have to integrate the posterior.
- Sometimes difficult or unfeasible. . .
- Solution: Markov-Chain Monte-Carlo

Bayes Theorem (revisited)

Bayes Theorem

$$\underbrace{p(H|y)}_{posterior} = \frac{\overbrace{p(y|H)}^{likelihood}\,\overbrace{p(H)}^{prior}}{p(y)}$$

Bayes Theorem (revisited)

Bayes Theorem

$$\underbrace{p(H|y)}_{\text{posterior}} = \frac{\overbrace{p(y|H)}^{\text{likelihood}}\overbrace{p(H)}^{\text{prior}}}{p(y)}$$

$$p(H|y) \propto p(y|H)p(H)$$

- $\Pr(y)$ can be ignored as constant for H.

Bayes Theorem (revisited)

Bayes Theorem

$$\underbrace{p(H|y)}_{\text{posterior}} = \frac{\overbrace{p(y|H)}^{\text{likelihood}}\,\overbrace{p(H)}^{\text{prior}}}{p(y)}$$

$$p(H|y) \propto p(y|H)p(H)$$

- $\Pr(y)$ can be ignored as constant for H.
- But: What is $\Pr(y)$? Probability of data?

Bayes Theorem (revisited)

Bayes Theorem

$$\underbrace{p(H|y)}_{\text{posterior}} = \frac{\overbrace{p(y|H)}^{\text{likelihood}}\,\overbrace{p(H)}^{\text{prior}}}{p(y)}$$

$$p(H|y) \propto p(y|H)p(H)$$

- $\Pr(y)$ can be ignored as constant for H.
- But: What is $\Pr(y)$? Probability of data?

If there are only two models: H_1 and H_0,

$$p(y|H_1)p(H_1) + p(y|H_0)p(H_0) = p(y)$$

Bayes Theorem (revisited)

Bayes Theorem

$$\underbrace{p(H|y)}_{\textit{posterior}} = \frac{\overbrace{p(y|H)}^{\textit{likelihood}}\,\overbrace{p(H)}^{\textit{prior}}}{p(y)}$$

$$p(H|y) \propto p(y|H)p(H)$$

- $\Pr(y)$ can be ignored as constant for H.
- But: What is $\Pr(y)$? Probability of data?

If there are only two models: H_1 and H_0,

$$p(y|H_1)p(H_1) + p(y|H_0)p(H_0) = p(y)$$

In general:

$$\sum_H p(y|H)p(H) = p(y)$$

Bayes Theorem (revisited)

The first trial

- Prior: $\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$
- The first trial: successful judgement.

Bayes Theorem (revisited)

The first trial

- Prior: $\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$
- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)}
\end{aligned}
$$

Bayes Theorem (revisited)

The first trial

- Prior: $\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$
- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\text{Likelihood}} \times \overbrace{0.5}^{\text{prior}}}{p(y)} = \frac{0.35}{p(y)}
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)}
$$

Bayes Theorem (revisited)

The first trial

- Prior: $\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$
- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\textit{Likelihood}} \times \overbrace{0.5}^{\textit{prior}}}{p(y)} = \frac{0.35}{p(y)} = \frac{0.35}{0.35 + 0.25} \approx 0.583
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)} = \frac{0.25}{0.35 + 0.25} \approx 0.417
$$

Bayes Theorem (revisited)

The first trial

- Prior: $\Pr(H_1) = 0.5$ and $\Pr(H_0) = 0.5$
- The first trial: successful judgement.

$$
\begin{aligned}
p(H_1|y) &= \frac{p(y|H_1)p(H_1)}{p(y)} \\
&= \frac{\overbrace{0.7^1 \cdot 0.3^0}^{\text{Likelihood}} \times \overbrace{0.5}^{\text{prior}}}{p(y)} = \frac{0.35}{p(y)} = \frac{0.35}{0.35 + 0.25} \approx 0.583
\end{aligned}
$$

$$
p(H_0|y) = \frac{0.5^1 \cdot 0.5^0 \times 0.5}{p(y)} = \frac{0.25}{p(y)} = \frac{0.25}{0.35 + 0.25} \approx 0.417
$$

- $p(y)$: normalizing constant ensuring $p(H_1|y) + p(H_0|y) = 1$.

Bayes Theorem (revisited)

Bayes Theorem

$$p(H|y) = \frac{p(y|H)p(H)}{p(y)}$$

Bayes Theorem (revisited)

Bayes Theorem

$$
\begin{aligned}
p(H|y) &= \frac{p(y|H)p(H)}{p(y)} \\
&= \frac{p(y|H)p(H)}{\sum_H p(y|H)p(H)}
\end{aligned}
$$

Bayes Theorem (revisited)

Bayes Theorem

$$
\begin{aligned}
p(H|y) &= \frac{p(y|H)p(H)}{p(y)} \\
&= \frac{p(y|H)p(H)}{\sum_H p(y|H)p(H)}
\end{aligned}
$$

In terms of densities:

$$
\begin{aligned}
f(\theta|y) &= \frac{f(y|\theta)f(\theta)}{f(y)} \\
&= \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta}
\end{aligned}
$$

Here again: integration is required!

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.
- The posterior also depends on your prior belief.

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.
- The posterior also depends on your prior belief.
- For a larger number of observations frequentists and Bayesian have similar results.

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.
- The posterior also depends on your prior belief.
- For a larger number of observations frequentists and Bayesian have similar results.
- Conjugacy of the beta distribution with the likelihood based on the binomial distribution.

Short summary

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.
- The posterior also depends on your prior belief.
- For a larger number of observations frequentists and Bayesian have similar results.
- Conjugacy of the beta distribution with the likelihood based on the binomial distribution.
- For more details: Susumu Shikano (2014), Bayesian estimation of regression models. Henning Best and Christof Wolf, eds. *The SAGE Handbook of Regression Analysis and Causal Inference*, Sage. p.31-54.

Back to the first question

Are people capable to estimate the ideological orientation based on facial traits?

Back to the first question

Are people capable to estimate the ideological orientation based on facial traits?

- 164 students of the University of Konstanz judged 35 pictures of student candidates unknown to them.

Back to the first question

Are people capable to estimate the ideological orientation based on facial traits?

- 164 students of the University of Konstanz judged 35 pictures of student candidates unknown to them.
- Average correct judgement: 60%.

Back to the first question

Are people capable to estimate the ideological orientation based on facial traits?

- 164 students of the University of Konstanz judged 35 pictures of student candidates unknown to them.
- Average correct judgement: 60%.
- But: A significant part of the successful judgements is due to the attractive halo effect.

Back to the first question

Are people capable to estimate the ideological orientation based on facial traits?

- 164 students of the University of Konstanz judged 35 pictures of student candidates unknown to them.
- Average correct judgement: 60%.
- But: A significant part of the successful judgements is due to the attractive halo effect.
- For more details: Michael Herrmann & Susumu Shikano (2016) Attractiveness and facial competence bias face-based inferences of candidate ideology. *Political Psychology*. 37(3): pp.401-17.

`http://onlinelibrary.wiley.com/doi/10.1111/pops.12256/abstract`