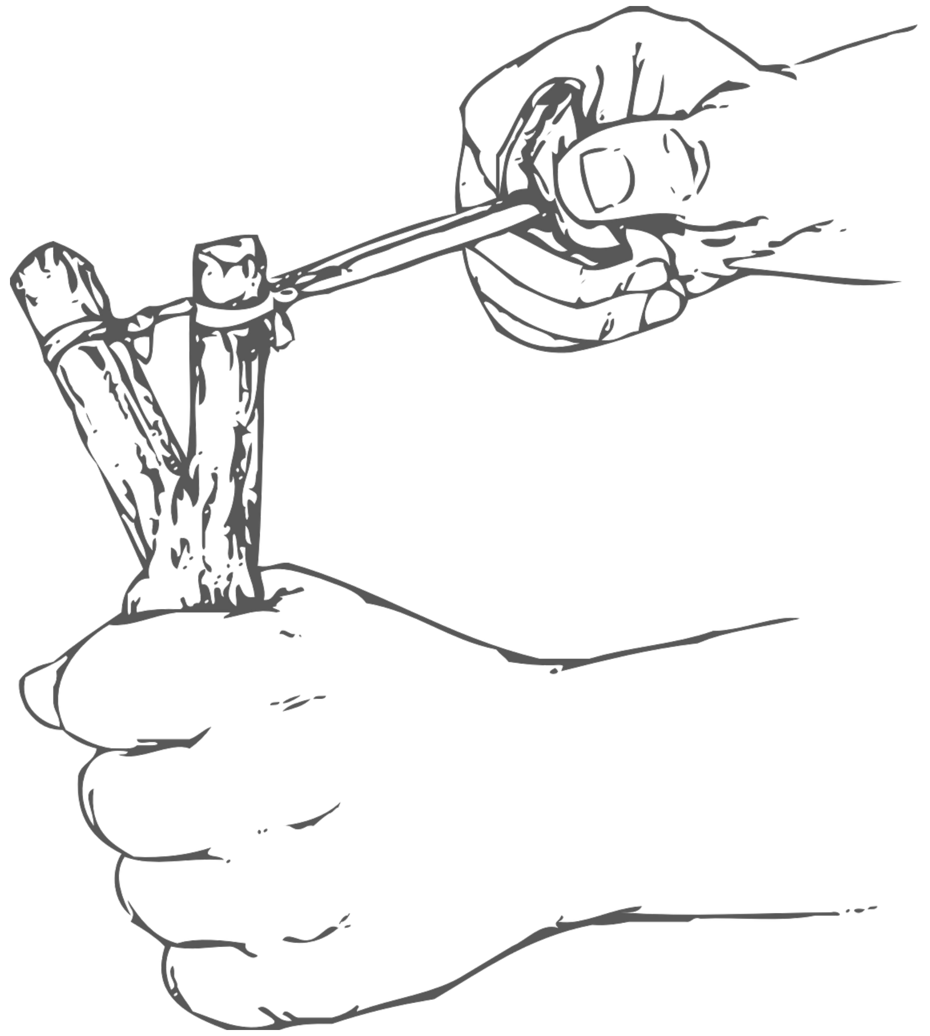
Bayesian data analysis: Theory & practice

Part 5a: Hypothesis testing

Michael Franke

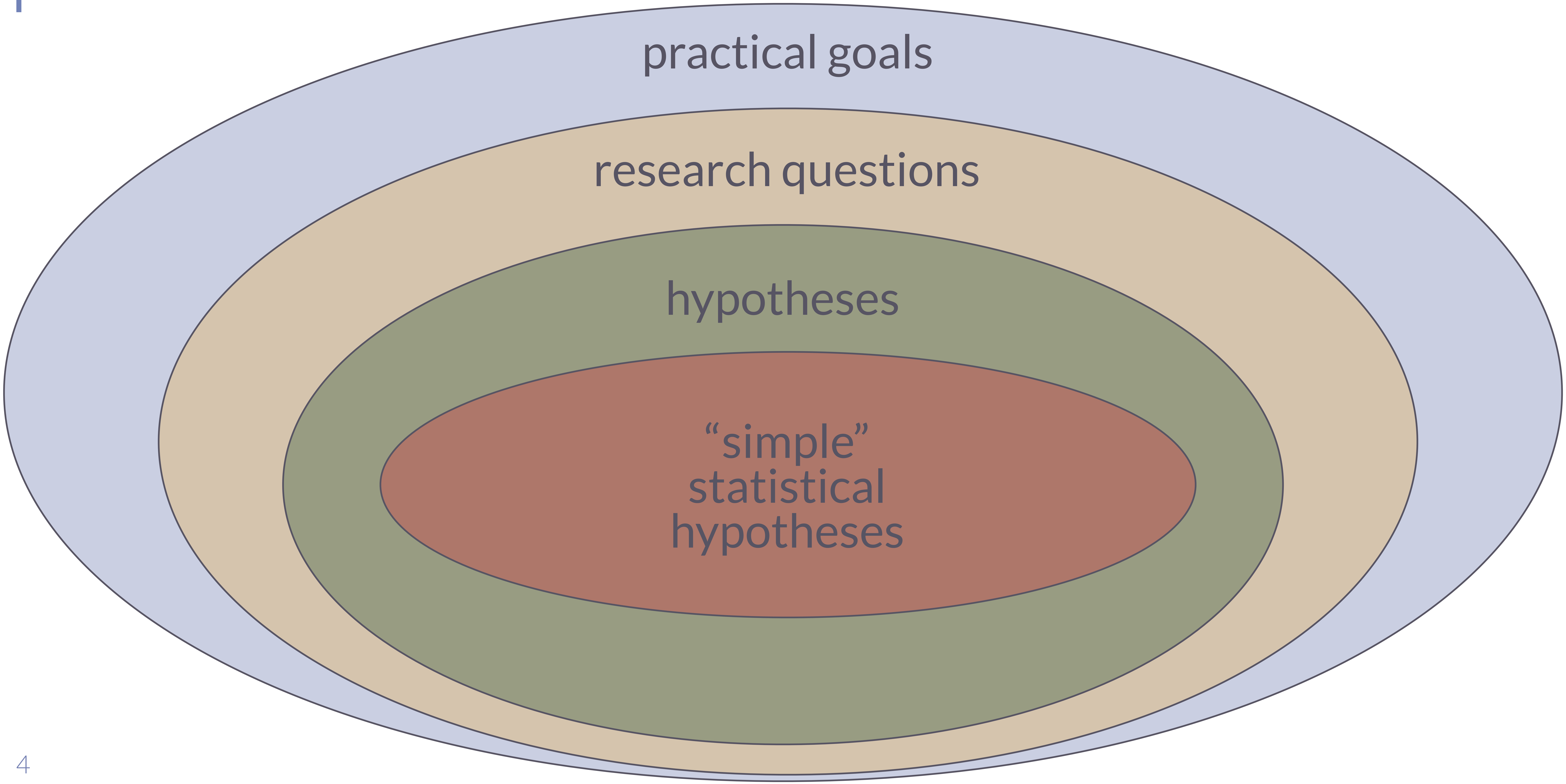
Main learning goals



1. reflect on the relation between research questions / goals and statistical hypotheses
2. get acquainted different methods of address hypotheses in BDA, based on:
   a. parameter estimation
   b. model checking
   c. model comparison
3. understand the conceptual and practical pros and cons of each
4. reflect on terminology for reporting empirical evidence

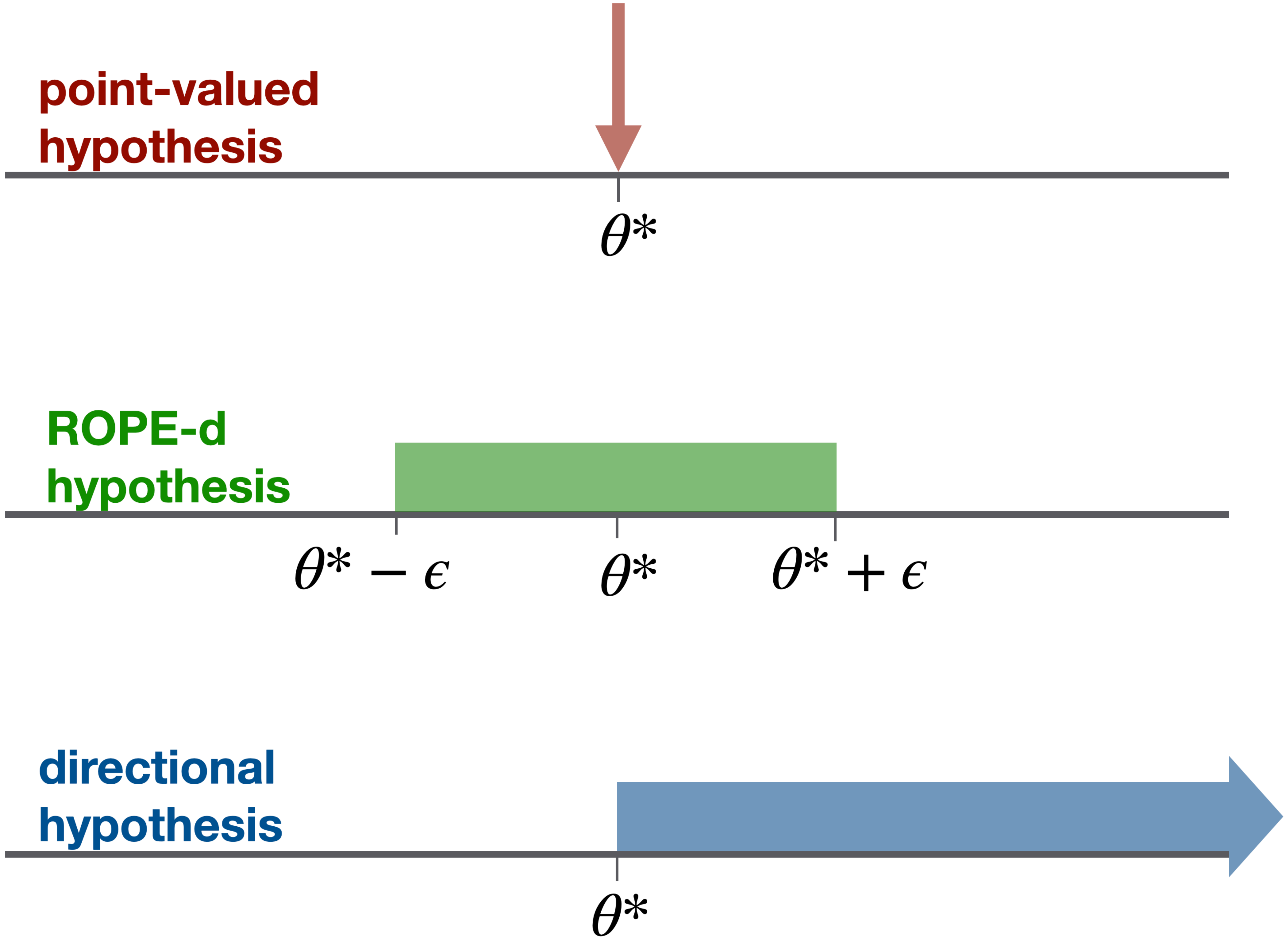
modeling goals, research questions, & statistical hypotheses

goals, questions, hypotheses



Types of statistical hypotheses

- point-valued $\theta = \theta^* \in \Theta$
- interval-valued $\theta \in I \subseteq \Theta$
  - ROPE-d
  - directional
- distributional $P \in \Delta(\Theta)$



read more here

approaches to testing statistical hypotheses

Three pillars of BDA

1. parameter estimation / inference

$$\underbrace{P(\theta \mid D)}_{\text{posterior}} \propto \underbrace{P(\theta)}_{\text{prior}} \times \underbrace{P(D \mid \theta)}_{\text{likelihood}}$$

2. predictions
   a. prior

$$P(D_{\text{pred}}) = \int P(\theta)\ P(D_{\text{pred}} \mid \theta)\ \mathrm{d}\theta$$

3. model comparison

$$\underbrace{\frac{P(M_1 \mid D)}{P(M_2 \mid D)}}_{\text{posterior odds}} = \underbrace{\frac{P(D \mid M_1)}{P(D \mid M_2)}}_{\text{Bayes factor}} \underbrace{\frac{P(M_1)}{P(M_2)}}_{\text{prior odds}}$$

Three ways to test a hypothesis

$\theta = \theta^*$

1. compute posterior, and check whether
   a. $P(\theta^* \mid D)$ high, and/or
   b. θ^* includes credible interval.
2. fix θ^* and perform prior / posterior predictive check (e.g., w/ likelihood as test statistic)
3. compare models with $\theta = \theta^*$ to another model

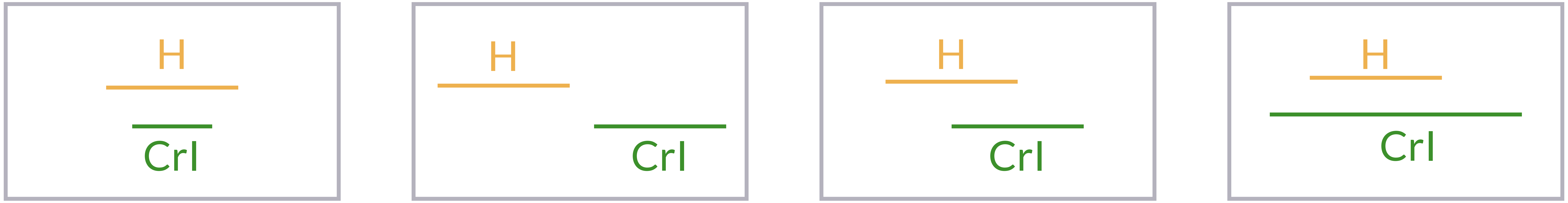
estimation-based testing

Estimation-based testing w/ HDIs

- interval-based hypothesis: $\theta \in I$
- posterior credible interval: $[l; u]$
- categorical approach:
  - **reason to accept** hypothesis I if $[l; u]$ is contained in I;
  - **reason to reject** hypothesis I if $[l; u]$ and I have no overlap;
  - **withhold judgement** otherwise.

text for preregistration / methods section

> We test hypothesis H by comparing a ROPE of $\pm$ 0.01 around the critical value $\theta = 0.5$ against a 95% credible interval of the posterior. We speak of *suggestive evidence in favor* of H, if the ROPE lies entirely inside the CredInt. […]

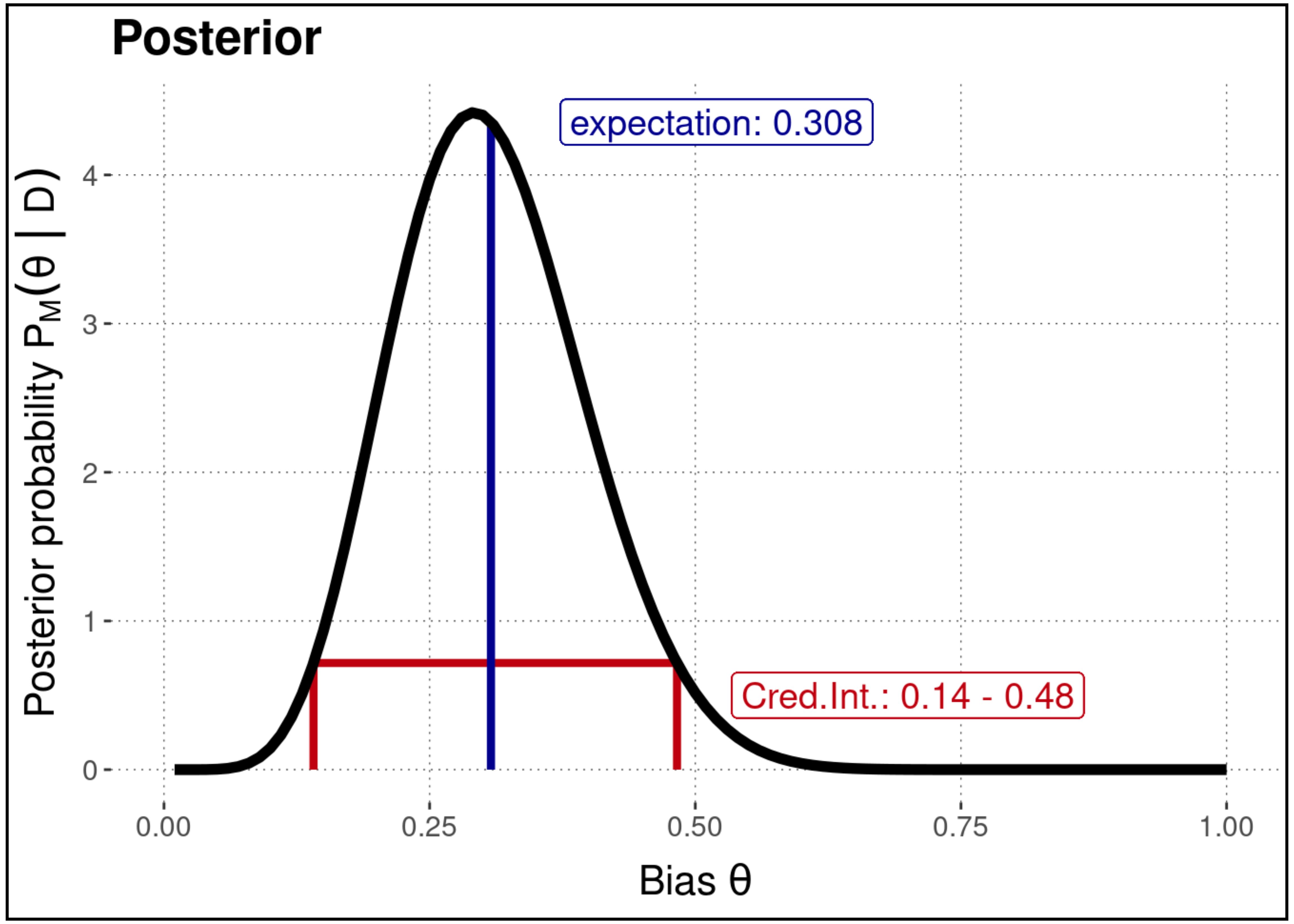


yes | no | maybe | maybe

read more here

Estimation-based testing with HDIs

example



text for results / discussion section

The 95% credible interval of the relevant parameter for hypothesis H is about [0.14; 0.48]. The defined ROPE [0.49; 0.51] lies entirely outside of that interval, so that, in line with preregistered / initially stated criteria, we interpret this as suggestive evidence *against* H.

Estimation-based testing w/ posterior probability

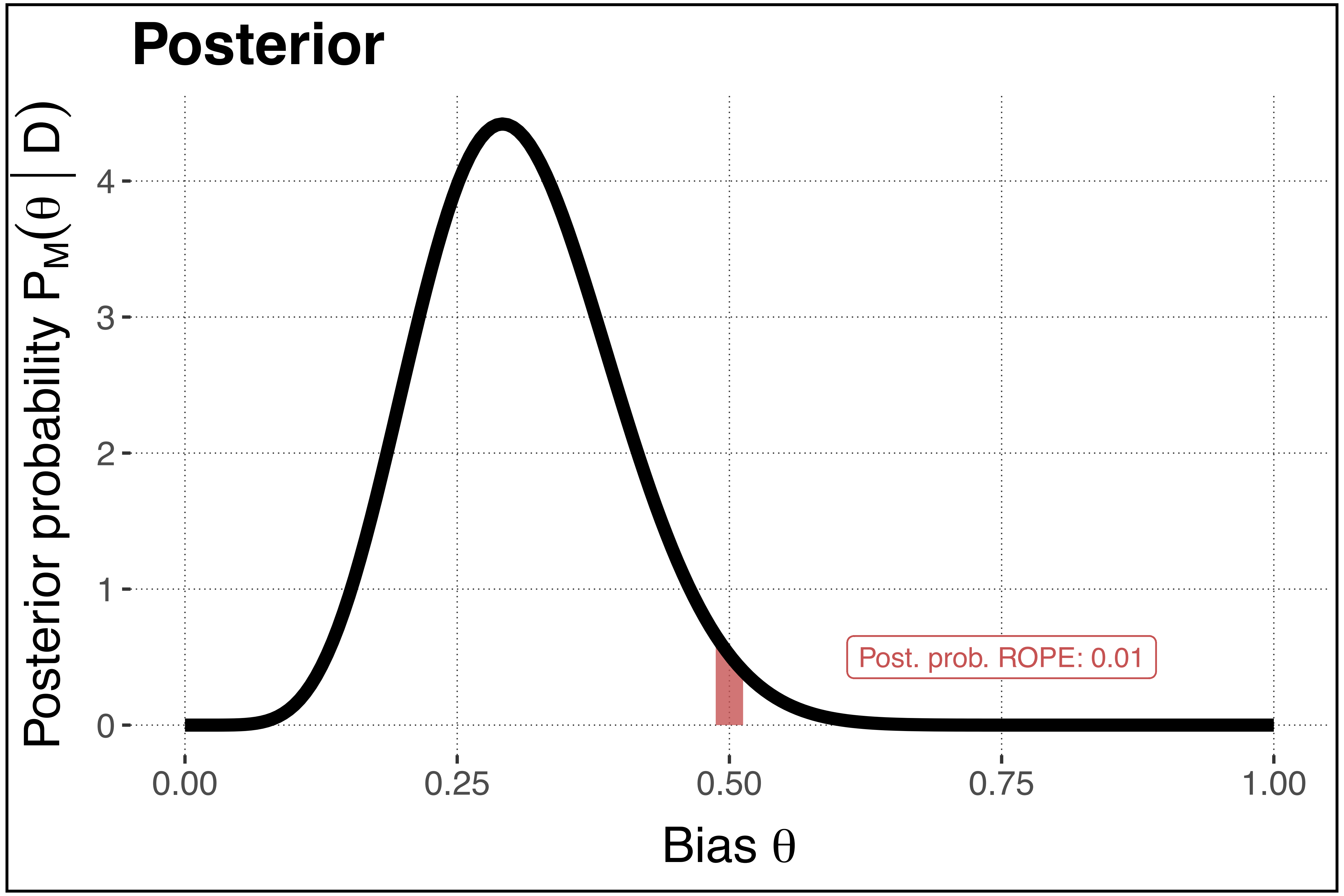
- interval-based hypothesis: $\theta \in I$
- **posterior probability** of I: $P(\theta \in I \mid D)$
- categorical approach:
  - **reason to accept** hypothesis I if $P(\theta \in I \mid D)$ is high;
  - **reason to reject** hypothesis I if $P(\theta \in I \mid D)$ is low;
  - **withhold judgement** otherwise.

text for preregistration / methods section

We test hypothesis H by considering a ROPE of $\pm$ 0.01 around the critical value $\theta = 0.5$. We speak of *suggestive evidence in favor* of H, if the posterior probability of the ROPE is at least 0.98. […]

Estimation-based testing w/ posterior probability

example



text for results / discussion section

The posterior probability of hypothesis H is about 0.01. In line with preregistered / initially stated criteria, we interpret this as suggestive evidence *against* H.

demo

hypothesis testing with Bayesian p-values

Bayesian predictive p-values

a generalization

- fix a model with $P(D \mid \theta)$ and $P(\theta)$
  - latter can be prior or posterior
  - gives prior / posterior predictive p-values
- $P_M(D)$ is the predictive distribution for model M
- Bayesian predictive p-value for observed data d_{obs}:

$$p(d_{\text{obs}}) = P_M\left(D \in \{d \mid P_M(d) \leq P_M(d_{\text{obs}})\}\right)$$

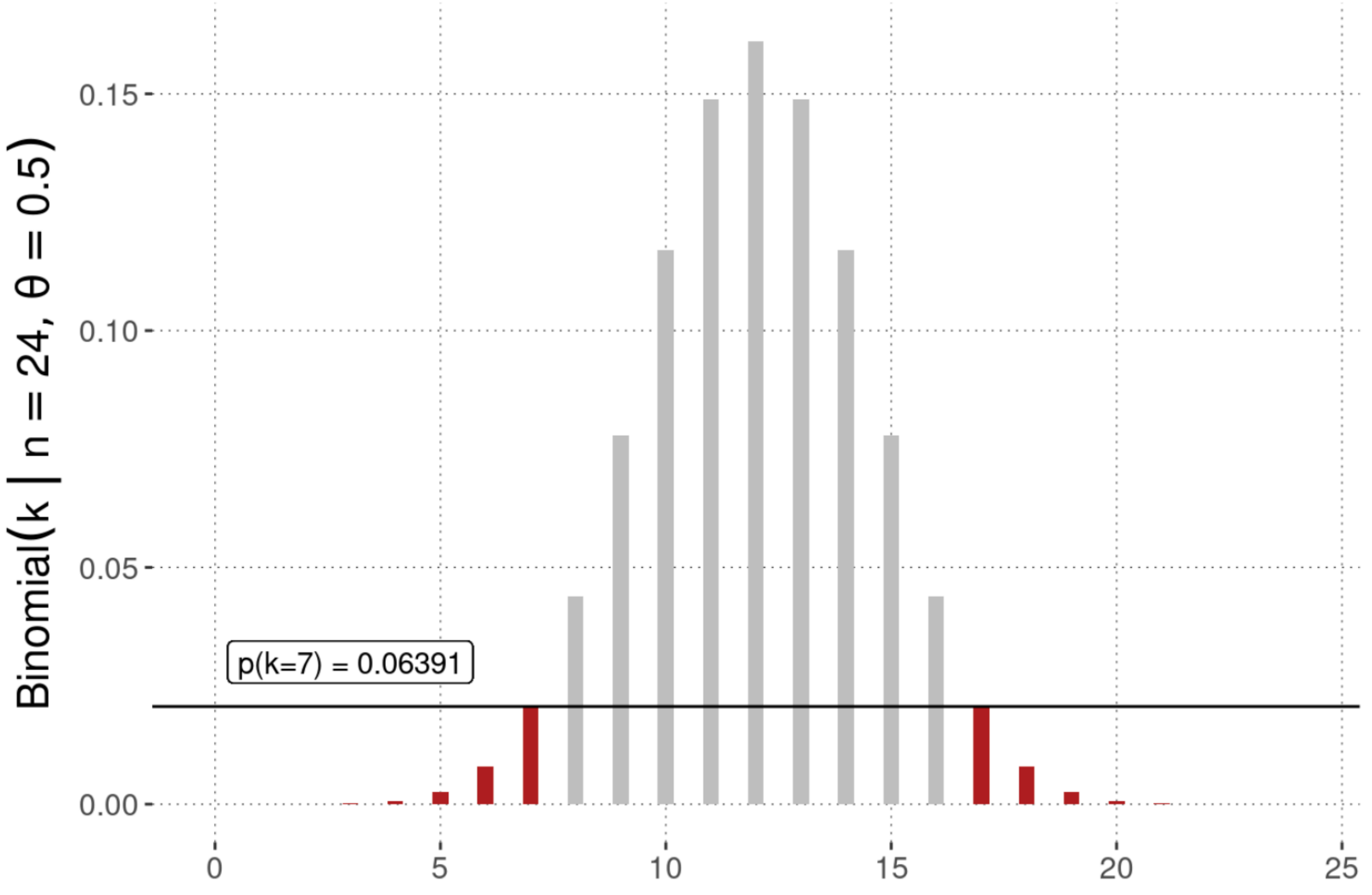
- approximated by sampling:

$$p(d_{\text{obs}}) \approx \frac{1}{n} \sum_{i=1}^{n} \left[P_M(d_i) \leq P_M(d_{\text{obs}})\right]$$

where $d_i \sim P_M(D)$ is a sample from the predictive distribution

Recap: frequentist p-values

$$p\left(D_{\text{obs}}\right) = P\left(T^{|H_0} \succeq^{H_{0,a}} t\left(D_{\text{obs}}\right)\right)$$



read more here

demo

Bayes factors for nested models

- Savage-Dickey method
- encompassing priors

Nested models

- suppose that there are n continuous parameters of interest $\theta = \langle \theta_1, \ldots, \theta_n \rangle$
- M_1 is a model defined by $P(\theta \mid M_1)$ & $P(D \mid \theta, M_1)$
- M_0 is **properly nested** under M_1 if:
  - M_0 assigns fixed values to some parameters $\theta_i = x_i, \ldots, \theta_n = x_n$
  - $\lim_{\theta_i \to x_i, \ldots, \theta_n \to x_n} P(\theta_1, \ldots, \theta_{i-1} \mid \theta_i, \ldots, \theta_n, M_1) = P(\theta_1, \ldots, \theta_{i-1} \mid M_0)$
  - $P(D \mid \theta_1, \ldots, \theta_{i-1}, M_0) = P(D \mid \theta_1, \ldots, \theta_{i-1}, \theta_i = x_i, \ldots, \theta_n = x_n, M_1)$

more here

Savage-Dickey method

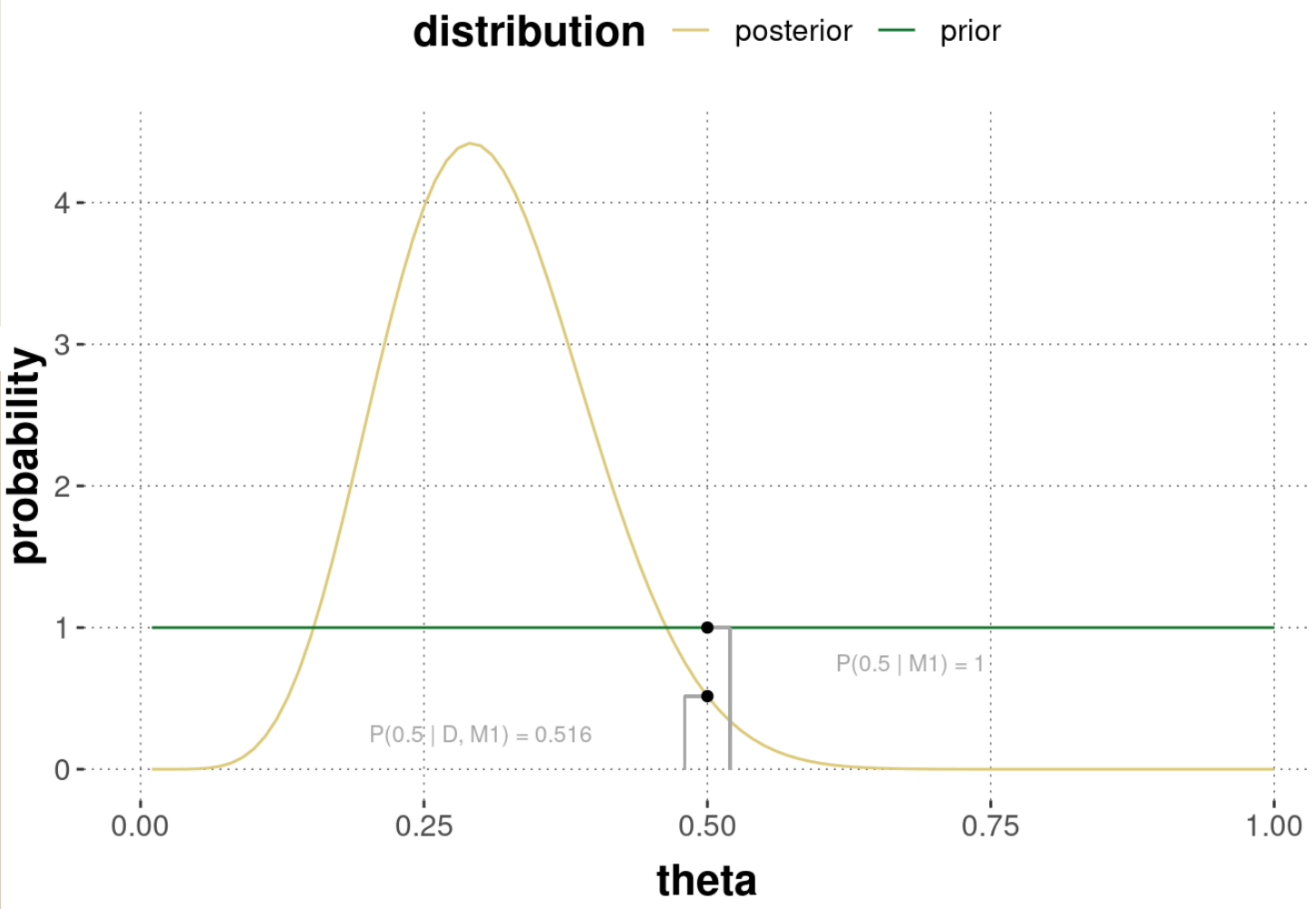
Theorem 11.1 (Savage-Dickey Bayes factors for nested models) *Let M_0 be properly nested under M_1 s.t. M_0 fixes $\theta_i = x_i, \dots, \theta_n = x_n$. The Bayes factor BF_{01} in favor of M_0 over M_1 is then given by the ratio of posterior probability to prior probability of the parameters $\theta_i = x_i, \dots, \theta_n = x_n$ from the point of view of the nesting model M_1:*

$$\mathrm{BF}_{01} = \frac{P(\theta_i = x_i, \dots, \theta_n = x_n \mid D, M_1)}{P(\theta_i = x_i, \dots, \theta_n = x_n \mid M_1)}$$

Proof. Let's assume that M_0 has parameters $\theta = \langle \phi, \psi \rangle$ with $\phi = \phi_0$, and that M_1 has parameters $\theta = \langle \phi, \psi \rangle$ with ϕ free to vary. If M_0 is properly nested under M_1, we know that $\lim_{\phi \to \phi_0} P(\psi \mid \phi, M_1) = P(\psi \mid M_0)$. We can then rewrite the marginal likelihood under M_0 as follows:

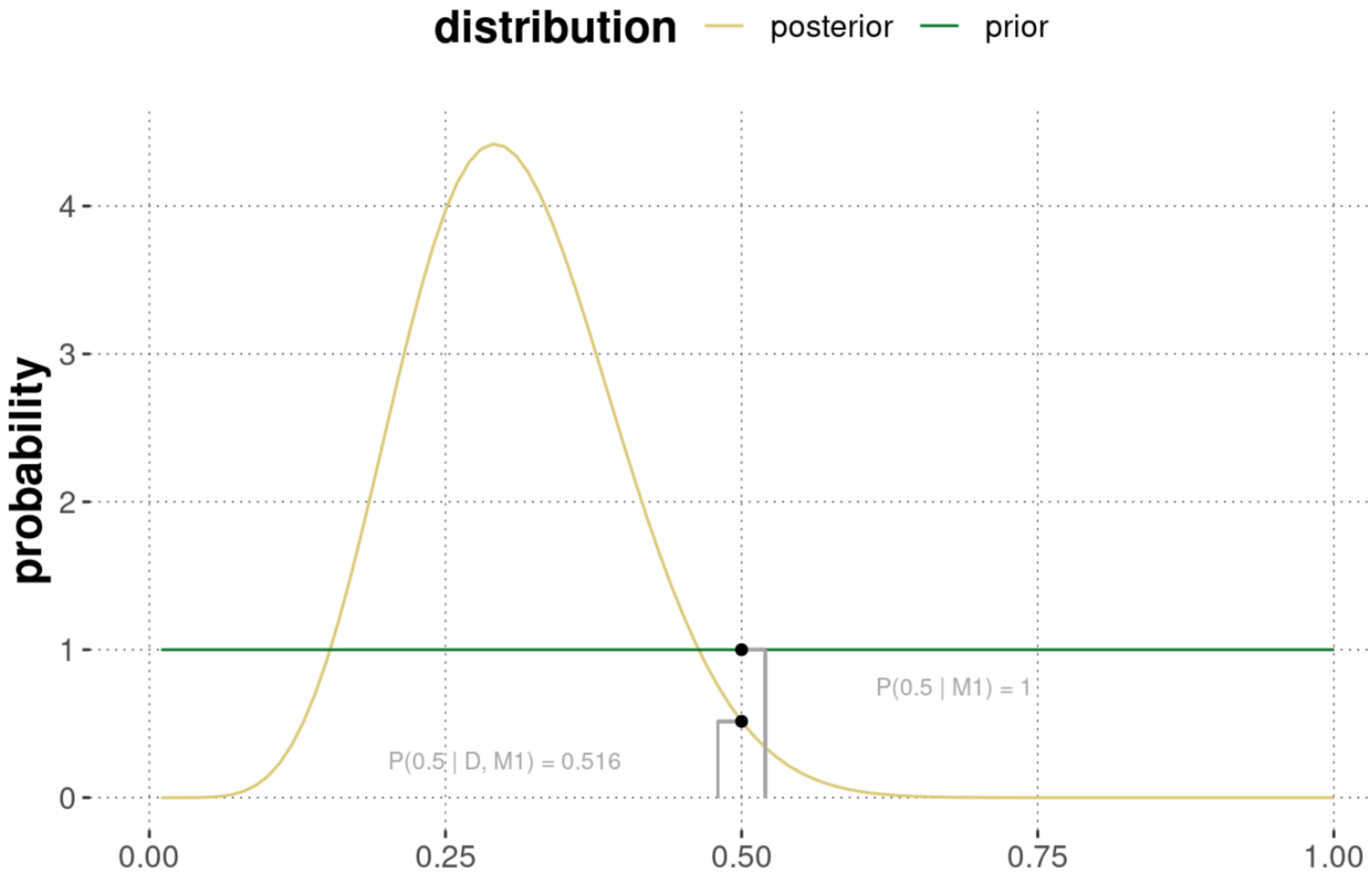
$$\begin{aligned}
P(D \mid M_0) &= \int P(D \mid \psi, M_0) P(\psi \mid M_0) \, \mathrm{d}\psi && [\text{marginalization}] \\
&= \int P(D \mid \psi, \phi = \phi_0, M_1) P(\psi \mid \phi = \phi_0, M_1) \, \mathrm{d}\psi && [\text{assumption of nesting}] \\
&= P(D \mid \phi = \phi_0, M_1) && [\text{marginalization}] \\
&= \frac{P(\phi = \phi_0 \mid D, M_1) P(D \mid M_1)}{P(\phi = \phi_0 \mid M_1)} && [\text{Bayes rule}]
\end{aligned}$$

The result follows if we divide by $P(D \mid M_1)$ on both sides of the equation. □



more here

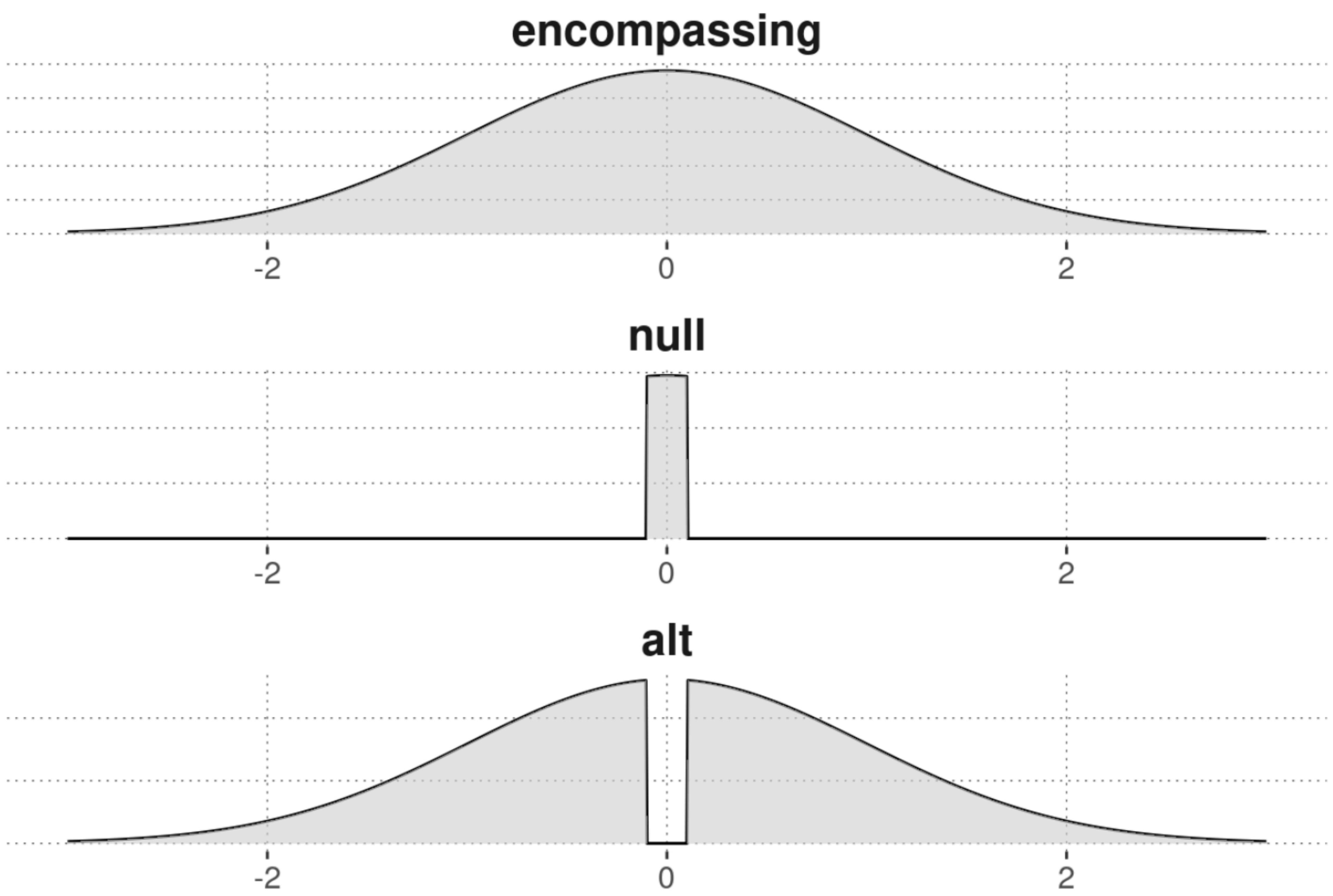
Savage-Dickey method



more here

Encompassing model

- target hypothesis is interval-based: $H_0: \theta \in I_0$
  - let I_1 be the complement of I_0
- an **encompassing model** M_e consists of:
  - likelihood $P(D \mid \omega, \theta, M_e)$
  - prior $P(\omega, \theta \mid M_e)$
- the **encompassed models** M_0 and M_1 share the likelihood function with M_e and have priors:
  - $P(\omega, \theta \mid M_i) = P(\omega, \theta \mid I_i, M_e)$



more here

generalized Savage-Dickey method

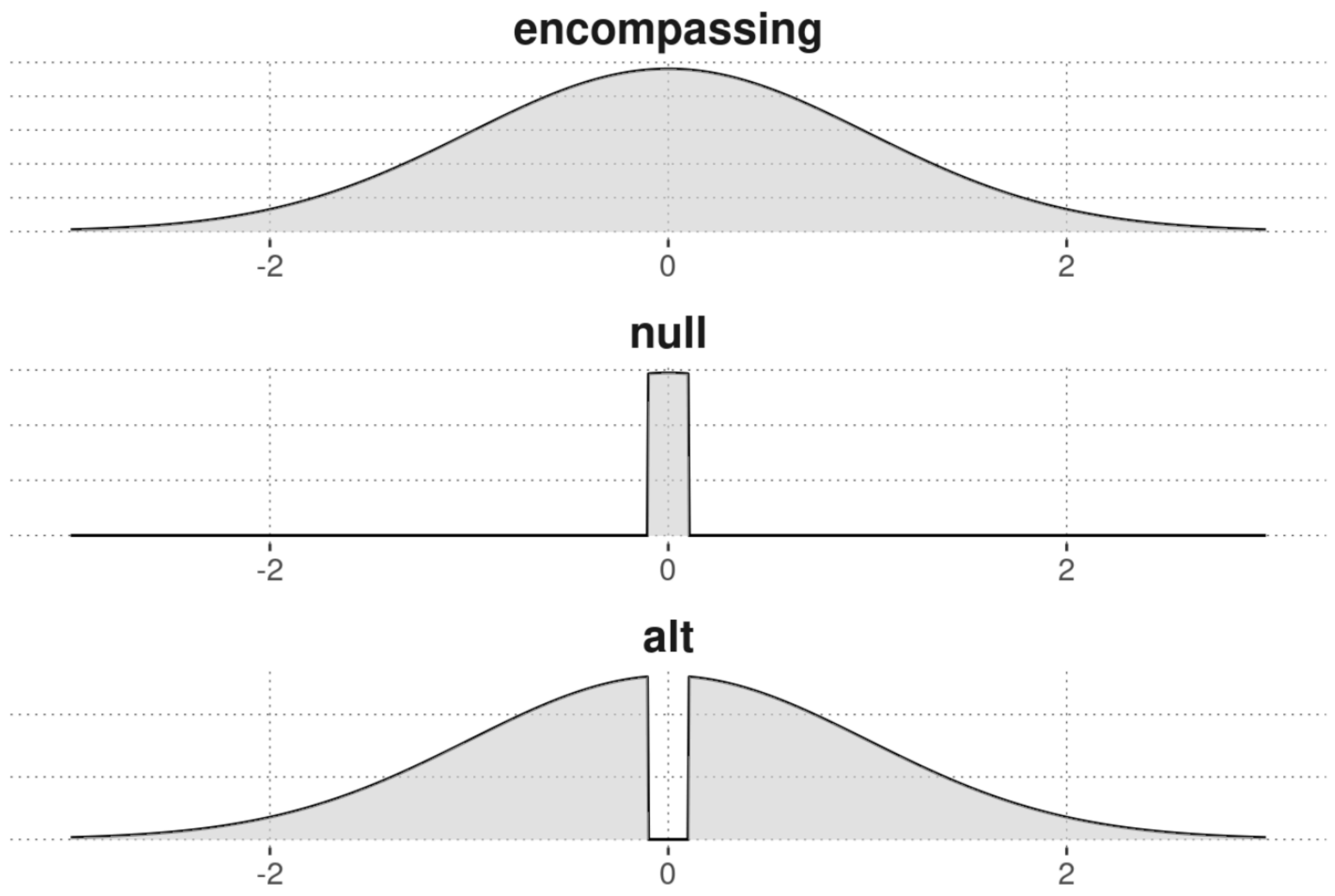
for encompassing models

Theorem 11.2 *The Bayes Factor in favor of nested model M_i over encompassing model M_e is:*

$$\mathrm{BF}_{ie} = \frac{P(\theta \in I_i \mid D, M_e)}{P(\theta \in I_i \mid M_e)}$$

Theorem 11.3 *The Bayes Factor in favor of model M_0 over alternative model M_1 is:*

$$\mathrm{BF}_{01} = \frac{P(\theta \in I_0 \mid D, M_e)}{P(\theta \in I_1 \mid D, M_e)} \frac{P(\theta \in I_1 \mid M_e)}{P(\theta \in I_0 \mid M_e)}$$



more here

LOO-based testing

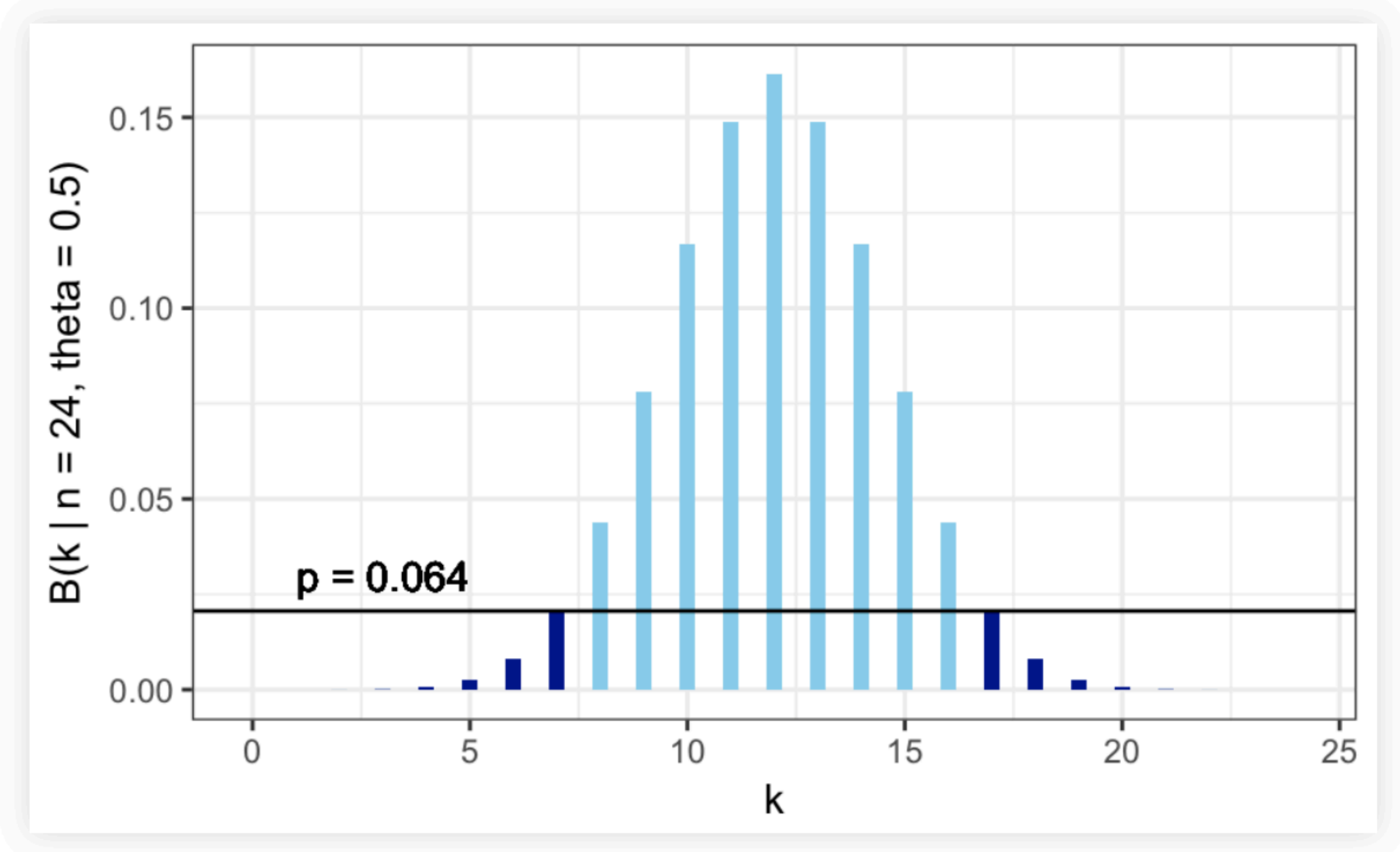
demo

Comparison

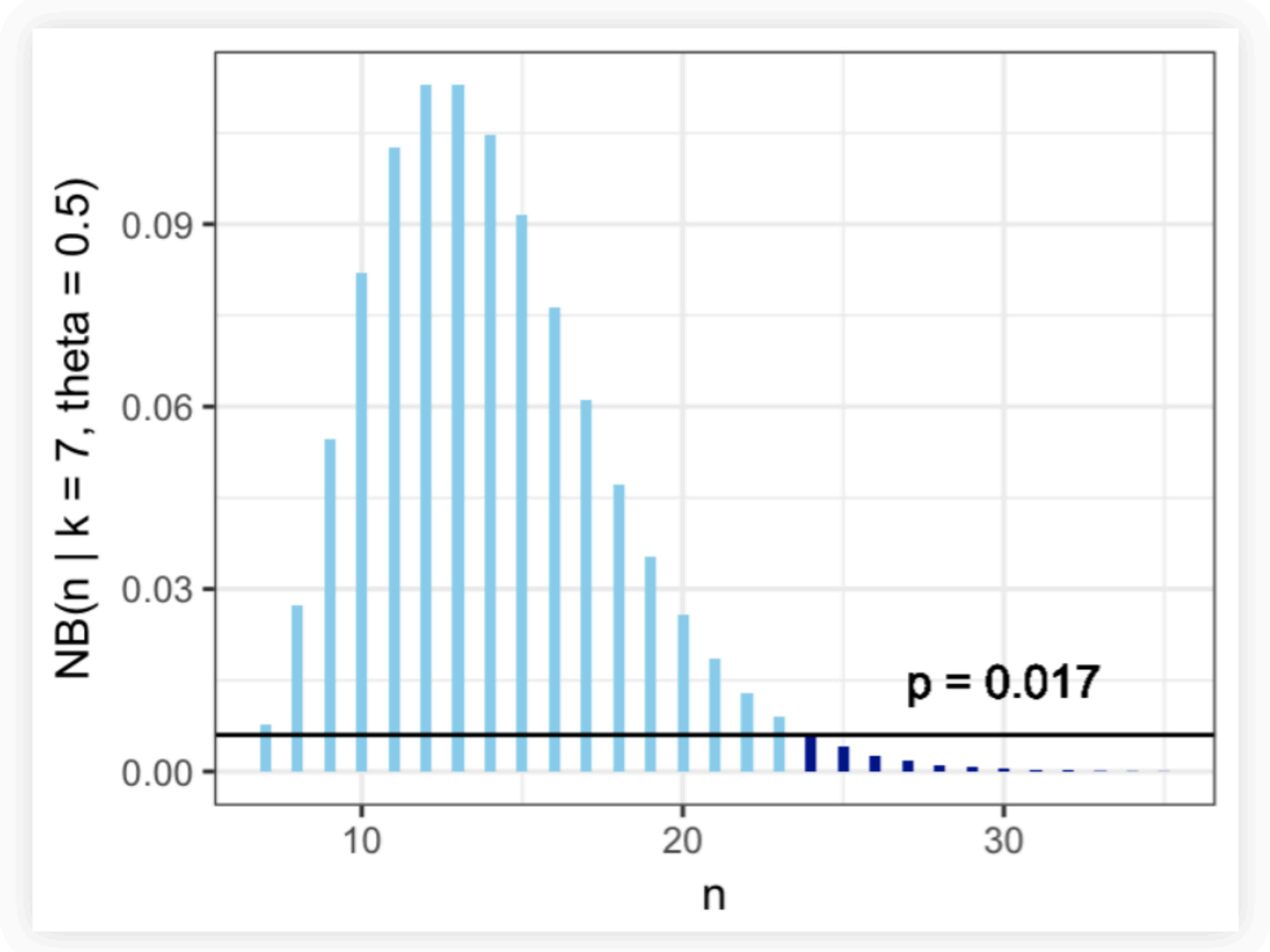
p-problems

- no evidence *for* the tested hypothesis
- dependence on sampling distribution

$$B(k; n = 24, \theta = 0.5) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$$



$$NB(n; k = 7, \theta = 0.5) = \frac{k}{n} \binom{n}{k} \theta^k (1 - \theta)^{n-k}$$



Jeffrey-Lindley paradox

clairvoyance?

data

```
k = 49581

N = 98451
```

frequentist p-value

```
binom.test(k, N)$p.value
```

```
## [1] 0.02364686
```

BF w/ Savage-Dickey
(alternative: flat prior)

```
dbeta(0.5, k + 1, N - k + 1)
```

```
## [1] 19.21139
```

Comparison of approaches

approach	method	computation	interpretation	measure	pro	con
estimation	Cred. Interval	easy	easy	reasonable to believe (categ.)	easy	dep. on prior
estimation	posterior prob.	easy	easy	level of credence (quant)	easy	dep. on prior
criticism	p-values	hard	hard	surprise (quant)	actual <u>test</u>	dep. on sampling distribution, only evidence *against* H
comparison	Bayes factors	hard	medium	relative strength of evidence (quant)	intuitive	dep. on alternative model & priors
comparison	LOO	medium	hard	post. predictive accuracy (quant)	cool and coming	unclear if actually a <u>test</u>