Bayes factor

Readings for today

- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. Psychonomic bulletin & review, 14(5), 779-804.

Topics

1. Problems with p-values

2. Bayes factors

Problems with p-values

Null hypothesis test statistics (NHTSs)

test statistic

$$p = P(Y \mid H_o = true) = P(T(X_{H_0}) \geq T(X_{obs}))$$

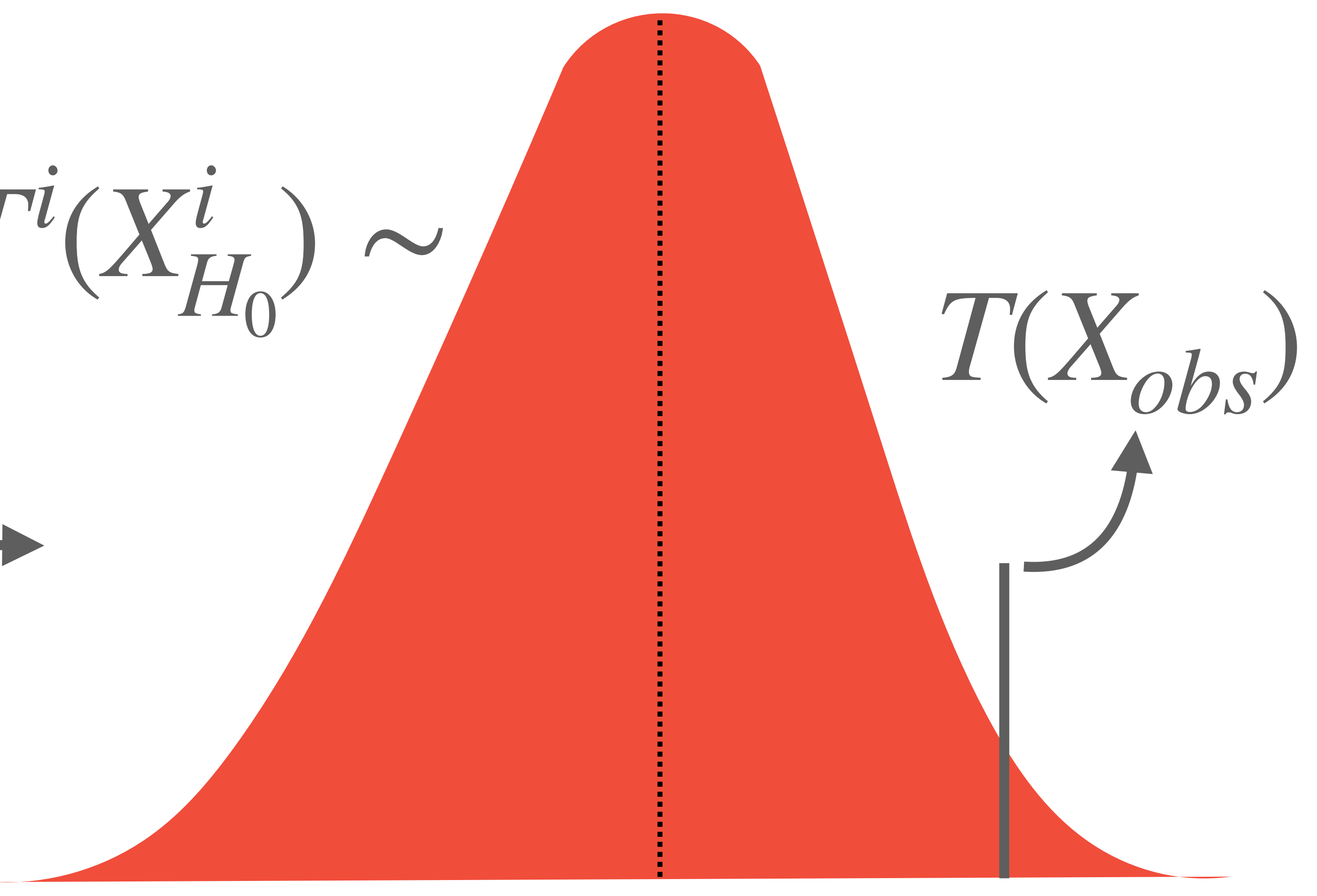
observed data

null hypothesis

predictor variables generated by H_0

predictor variables you observe

$$H_0 \rightarrow \left.\begin{matrix} T^1(X^1_{H_0}) \\ \vdots \\ T^n(X^n_{H_0}) \end{matrix}\right\} \longrightarrow T^i(X^i_{H_0}) \sim$$



(Wasserstein et al. 2019)

Problems with p-values

Fundamental limitations:

1. Depend on unobserved data.
2. Depend on subjective intentions.
3. Do not quantify statistical evidence.
4. Poorly understood.

probability that
the null *exists*.

$$p = P(Y \mid H_o = true)$$

(Wasserstein et al. 2019)

Ideal measure of hypotheses

Ideal traits of hypothesis evaluation:

1. Depends only on observed data.
2. Is independent of subjective intentions.
3. Is a measure of evidence, not existence.
4. Is easy to interpret.
5. Is objective.

Want

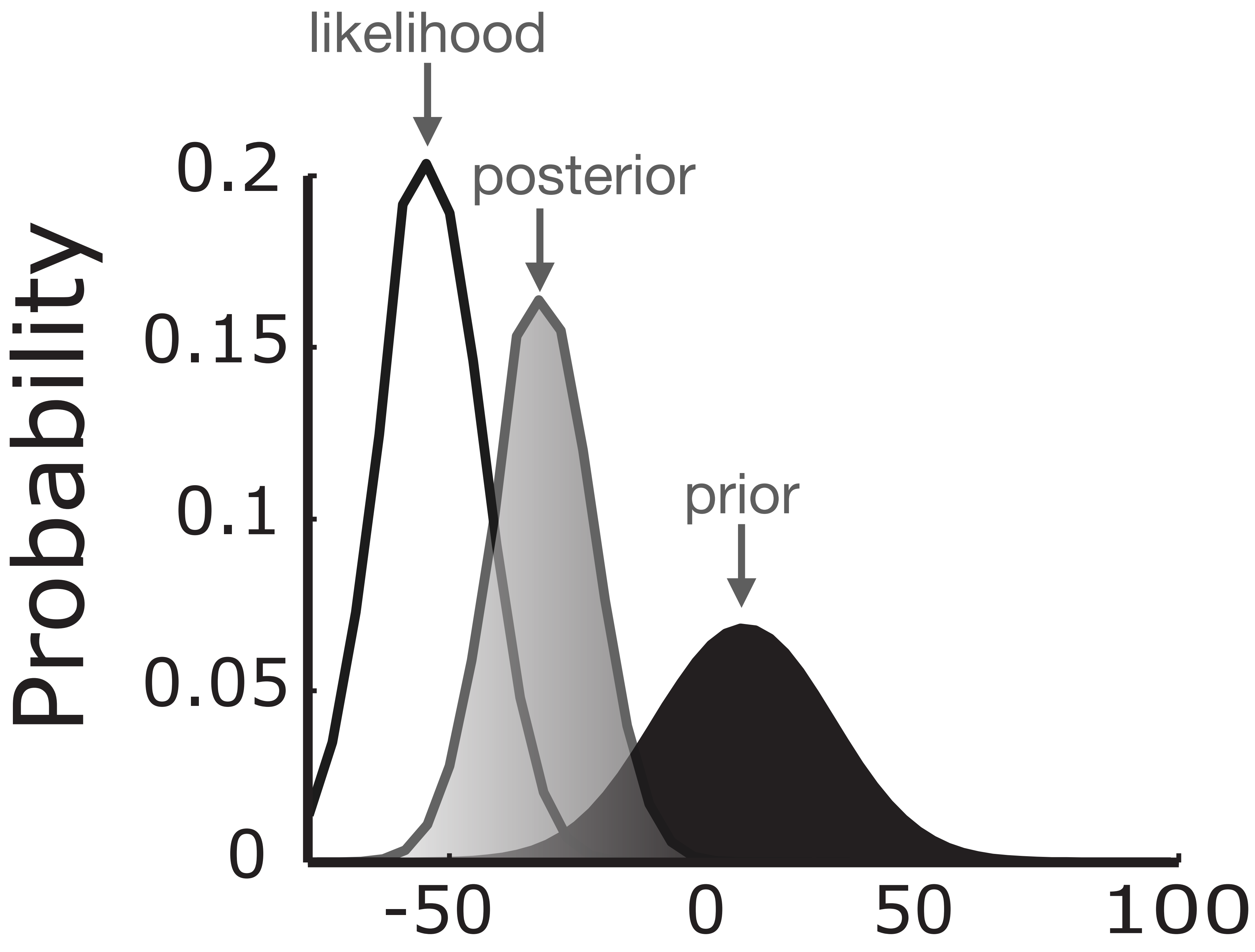
$$\overbrace{P(H_o = true \mid Y)}^{\text{evidence}} \text{ not } \overbrace{P(Y \mid H_o = true)}^{\text{existence}}$$

(Wagenmakers 2007)

Bayes factor

Bayes Theorem

$$\underbrace{P(H_i \mid Y)}_{\text{posterior}} = \frac{\overbrace{P(Y \mid H_i)}^{\text{likelihood}} \; \overbrace{P(H_i)}^{\text{prior}}}{\underbrace{P(Y)}_{\text{marginal}}}$$



(Wagenmakers 2007)

Example: (Gaussian) linear model

model: $$\hat{f}(X) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j X_j$$

likelihood: $$P(Y\,|\,\hat{f}(X)) = P(Y\,|\,X, \hat{\beta}) = \sum_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \beta_j x_i)^2}{2\sigma^2}}$$

prior: $$P(\hat{f}(X)) = P(X, \hat{\beta}) = \underbrace{U(-\infty, +\infty)}_{\text{unknown}} \text{ or } \underbrace{N(\mu, \sigma)}_{\text{known}}$$

in log form

posterior: $$P(\hat{f}(X)\,|\,Y) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_i x_{i,j})^2}{2\sigma}$$

(Wagenmakers 2007)

Bayes Factor

What is the ratio of evidence for two hypotheses?

$$BF_{01} = \frac{P(H_0 \mid Y)}{P(H_1 \mid Y)} = \frac{P(Y \mid H_0)}{P(Y \mid H_1)} \cdot \frac{P(H_0)}{P(H_1)}$$

Bayes Factor <u>for</u> H_0 (against H_1)

- Determine the <u>relative</u> evidence for one hypothesis against the other.
- BF_{ij} identifies whether the observed data are more likely to arise from hypothesis i (H_i) than from hypothesis j (H_j).

(Wagenmakers 2007)

Example: (Gaussian) linear model

$$\underline{H_1}: \quad \hat{f}(X) = \hat{\beta}_0 + \overbrace{\hat{\beta}_1 X_1}^{confound} + \overbrace{\hat{\beta}_2 X_2}^{target}$$

$$\underline{H_0}: \quad \hat{f}(X) = \hat{\beta}_0 + \underbrace{\hat{\beta}_1 X_1}_{confound}$$

<u>Bayesian Information Criteria (BIC)</u>

$$BIC(H_i) = p_i \log(n) - 2\log(L_i)$$

\# parameters

$$L_i = \frac{RSS_i}{TSS}$$

<u>Difference in BICs</u>

$$\Delta BIC_{10} = BIC(H_1) - BIC(H_0)$$

$$= n\log(\frac{RSS_1}{RSS_0}) + (p_1 - p_0)\log(n)$$

$H_1 - H_0$

<u>Bayes Factor for H_0</u>

$$BF_{01} \propto \frac{P(Y|H_0)}{P(Y|H_1)} = e^{\frac{\Delta BIC_{10}}{2}}$$

(Wagenmakers 2007)

Bayes Factor

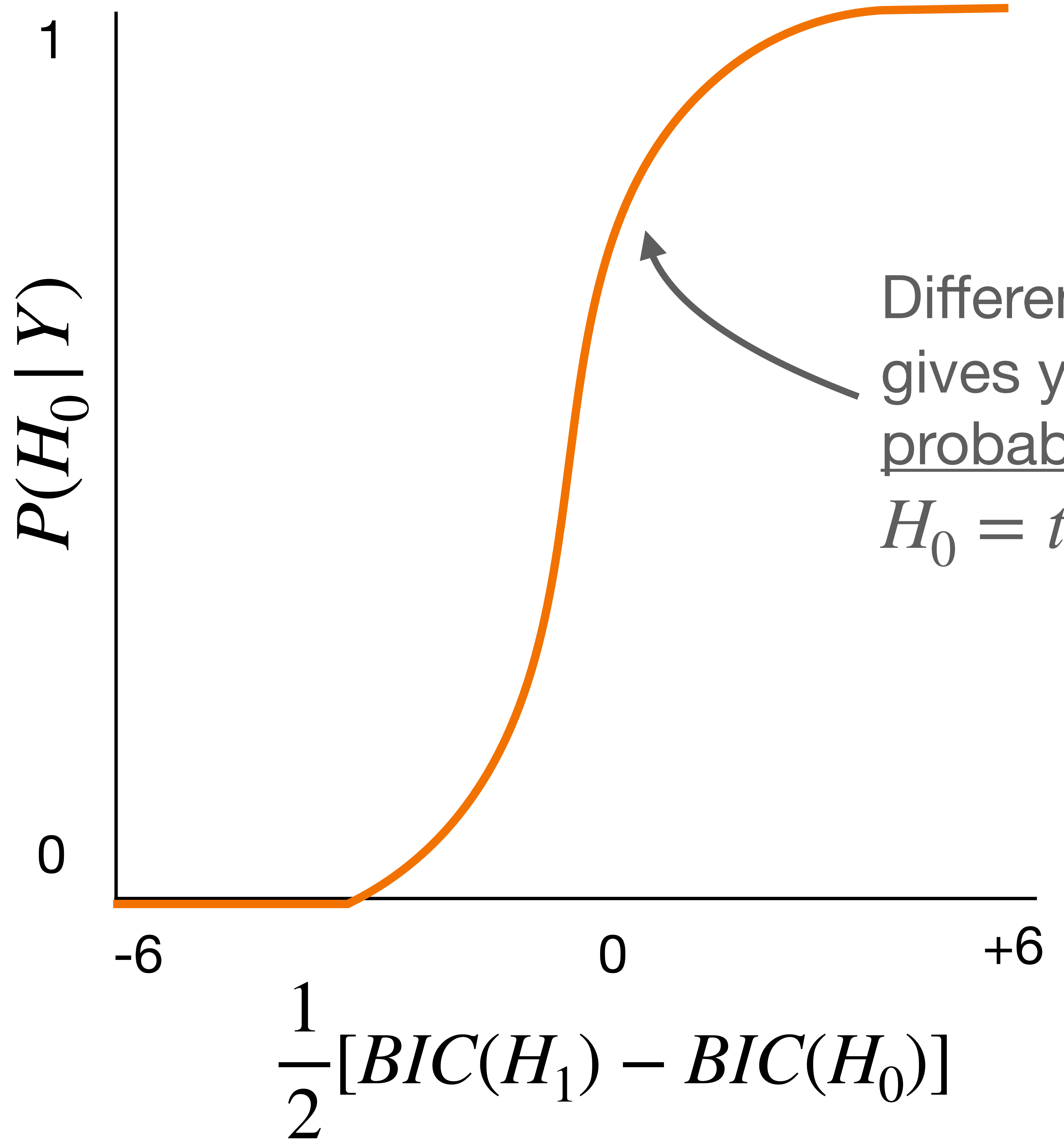
General formulation of BFs: Evaluate k-many alternative models

One against many

$$P(H_i\,|\,Y) = \frac{e^{-\frac{BIC(H_i)}{2}}}{\sum_{j=0}^{k\neg i} e^{-\frac{BIC(H_j)}{2}}}$$

H_0 alone

$$P(H_0\,|\,Y) = \frac{1}{1 + e^{-\frac{\Delta BIC_{10}}{2}}}$$



Difference in $BICs$ gives you the precise probability of $H_0 = true$

(Wagenmakers 2007)

Inferring from BFs

No equivalent of $p < 0.05$ for BFs, so have to make inferential heuristics based on the strength of evidence.

BF_{01}	$P(H_0 \mid Y)$	Evidence
1-3	0.50-0.75	weak
3-20	0.75-0.95	positive
20-150	0.95-0.99	stong
>150	>0.99	very strong

(Wagenmakers 2007)

Take home message

- Bayes Factors offer the ability to directly estimate the evidence for one hypothesis over another. This is incredibly useful for making inferences for or against the null hypothesis itself.