

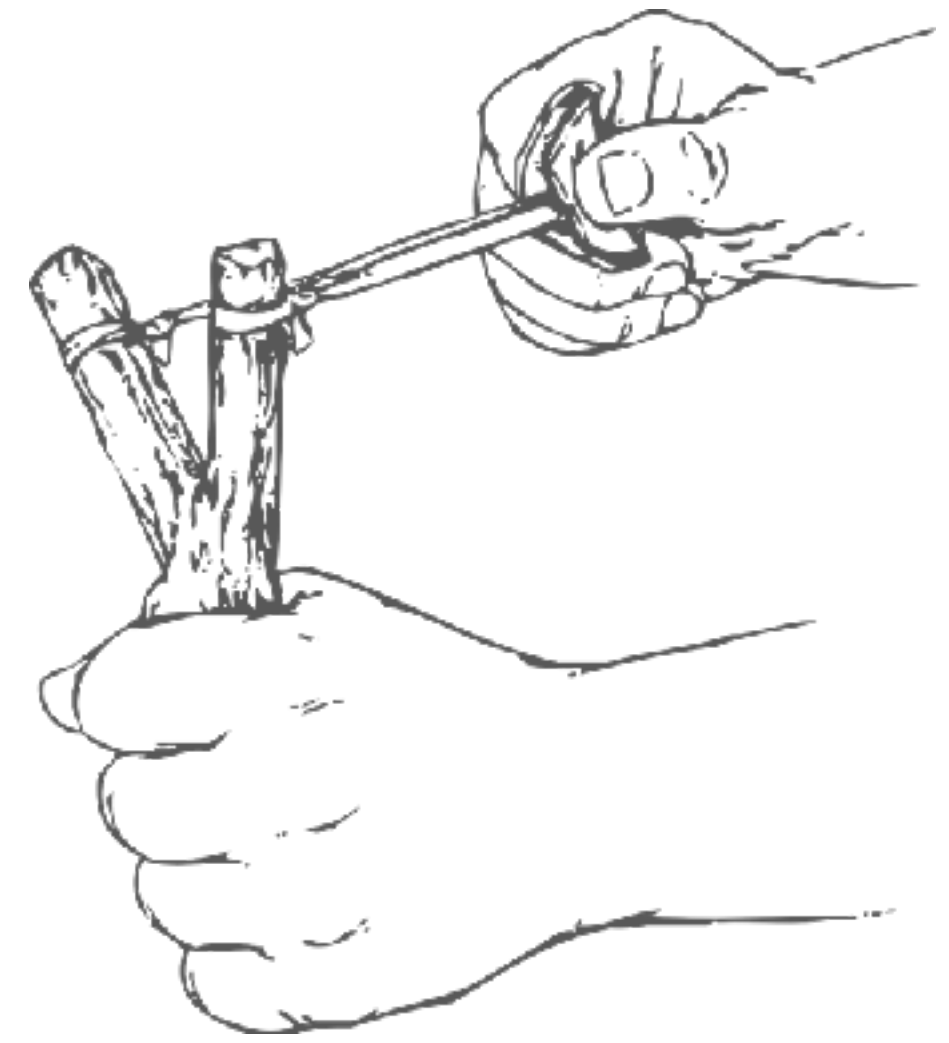
Bayesian data analysis: Theory & practice

Part 4b: Model checking

Michael Franke

Main learning goals

1. understand the role of model checking in statistical inquiry
 - a. assessing implications of priors
 - b. inspecting posterior predictives
2. apply common methods of posterior predictive checking
 - a. visual
 - b. Bayesian p -values



Three pillars of BDA

1. parameter estimation / inference [which parameter values are credible given data and model?]

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

2. predictions [which future data observations are likely given my model?]

a. prior

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

b. posterior

$$P(D_{\text{pred}} | D_{\text{obs}}) = \int P(\theta | D_{\text{obs}}) P(D_{\text{pred}} | \theta) d\theta$$

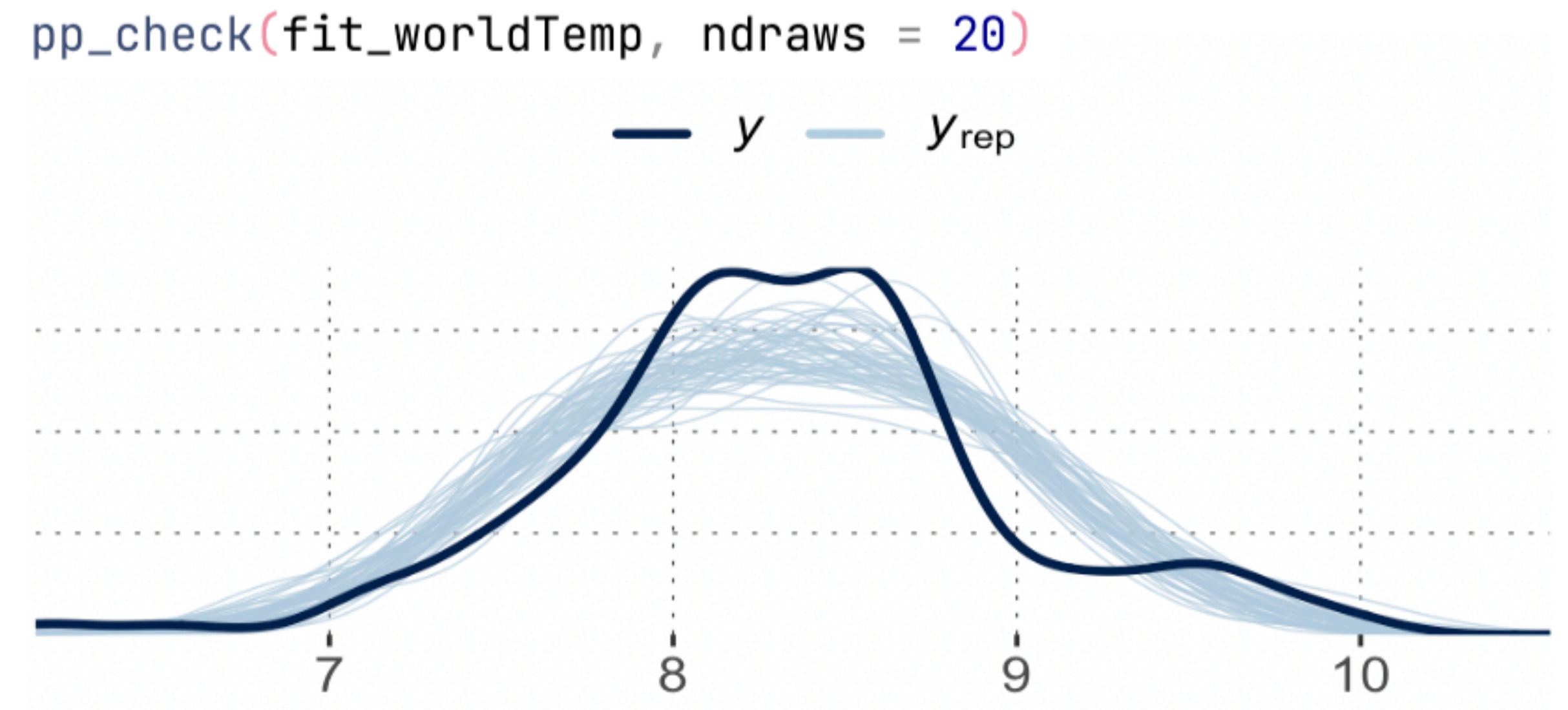
3. model comparison [which model of two models is more likely to have generated the data?]

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

Visual posterior predictive checks

for world-temperature data

- ▶ black line:
 - distribution of observed temperature
- ▶ each of the 50 blue lines:
 - distribution of temperatures predicted for same years given a sample from the posterior



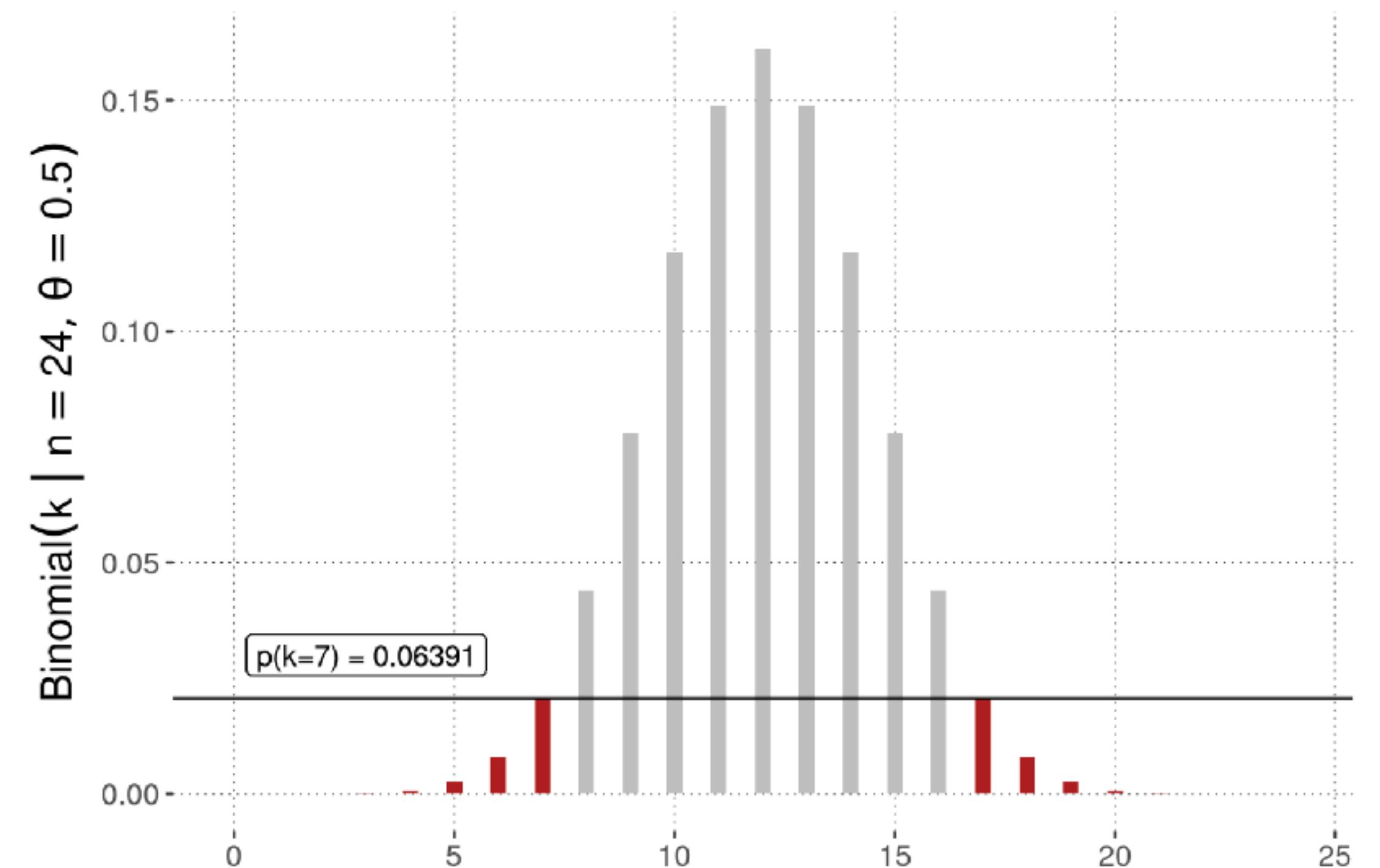
Recap: p -values

Monte Carlo approximation

- ▶ fix null hypothesis $\theta = \theta^*$
- ▶ derive **sampling distribution** $P(D | \theta^*)$
 - how likely is each possible observation under θ^*
- ▶ fix a **test statistic** $t(D)$
 - real number measuring some relevant aspect of D
- ▶ consider observed data D_{obs}
- ▶ p -value from MC simulation:
 - sample $d_1, \dots, d_n \sim P(D | \theta^*)$
 - calculate:

$$p(D_{\text{obs}}) \approx \frac{1}{n} \sum_{i=1}^n [P(d_i | \theta^*) \leq P(D_{\text{obs}} | \theta^*)]$$

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



Bayesian p -values

Monte Carlo approximation

- ▶ fix a model with $P(D \mid \theta)$ and $P(\theta)$
 - latter can be prior or posterior
 - gives prior / posterior predictive p -values
- ▶ gives **predictive distribution** $P_M(D)$
- ▶ fix a **test statistic** $t(D)$
 - real number measuring some relevant aspect of D
- ▶ consider observed data D_{obs}
- ▶ p -value from MC simulation:
 - sample $d_1, \dots, d_n \sim P_M(D)$
 - calculate:

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

demo



visual PPCs and Bayesian p