

Is the p value dead? Frequentist vs. Bayesian inference

5.1 Introduction: Frequentist vs. Bayesian inference

5.2 General Methodics

5.3 Discrete Quantities and Observations

5.3.1 Example: Covid-19 Tests

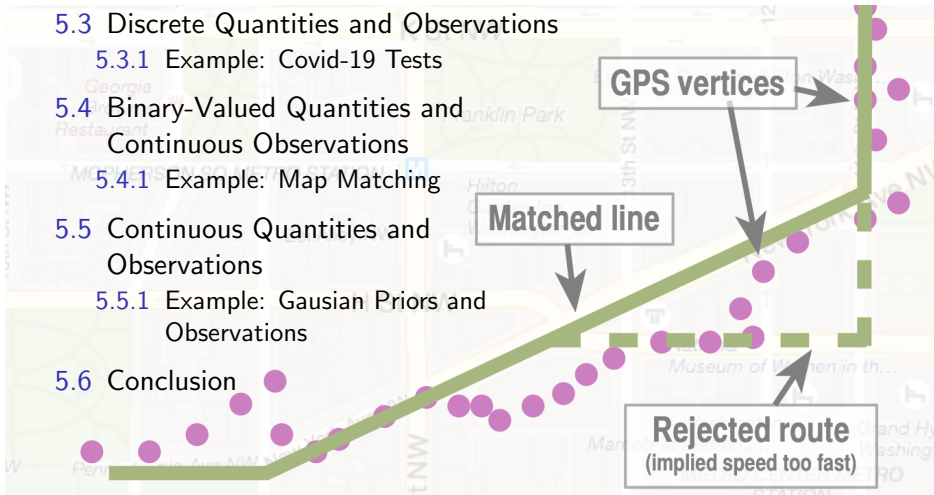
5.4 Binary-Valued Quantities and Continuous Observations

5.4.1 Example: Map Matching

5.5 Continuous Quantities and Observations

5.5.1 Example: Gaussian Priors and Observations

5.6 Conclusion



5.1 Introduction: Frequentist vs. Bayesian inference

- ▶ The classic **frequentist's** approach calculates the probability that the test function T is further away from H_0 , (in the extreme range E_{data}) than the data realisation provided H_0 is marginally true:

$$p = P(T \in E_{\text{data}} | H_0^*) \geq P(T \in E_{\text{data}} | H_0)$$

- ▶ The **Bayesian inference** tries to calculate what is actually interesting: The probability of H_0 given the data.
- ▶ If the unconditional or **a-priori probabilities** were known, this is easy using **Bayes' theorem** (abbreviating $T \in E_{\text{data}}$ as E_{data})

$$P(H_0 | E_{\text{data}}) = \frac{P(E_{\text{data}} | H_0) P(H_0)}{P(E_{\text{data}})} \leq p \frac{P(H_0)}{P(E_{\text{data}})}$$

- ▶ For real-valued parameters, this obviously makes only sense for interval null hypotheses since, for a point null hypothesis, we have exactly $P(H_0 | E_{\text{data}}) = P(H_0) = 0$.

5.2 General Idea

- ▶ Principle: Update the a-priori probability $P(H_0)$ of some event H_0 (in particular, a null hypothesis) based on an observation B , e.g., $B : \hat{\beta} = b$ or $B : \hat{\beta} \in [b - \delta/2, b + \delta/2]$ with some small δ
- ▶ Example: H_0 : “tomorrow is nice weather”
 - ▶ $P(H_0)$: a-priori probability before hearing the weather forecast (or the general probability based on climate tables)
 - ▶ B : tomorrow's weather forecast $B \in \{\text{will be nice, not nice}\}$
 - ▶ $P(H_0|B)$: a-posteriori probability after hearing the forecast
- ▶ Relation to classical frequentist's statistics: Known are some observation B and conditional probability $P(B|H_0)$ that can be expressed in terms of p . Want $P(H_0|B)$
- ▶ Four scaling possibilities
 - (i) discrete β and $\hat{\beta}$ (e.g., Covid-19 test)
 - (ii) discrete β and continuous $\hat{\beta}$ (e.g., map-matching)
 - (iii) continuous β , discrete observation (H_0 rejected or not)
 - (iv) continuous sought-after quantity β (continuous H_0) and continuous observation $\hat{\beta}$ (e.g., regression models)

5.3 Bayesian Inference for Discrete Quantities and Observations

The classical textbook case are binary variables (values “true”, “false”, each set has one element; generalisations easy):

$$H_0 : \beta = \text{true}, \quad \bar{H}_0 : \beta = \text{false}, \quad B : \hat{\beta} = \text{true}; \quad \bar{B} : \hat{\beta} = \text{false}$$

$$P(H_0|B) = \frac{P(B|H_0)P(H_0)}{P(B)}$$

Example: Covid-19 tests

- ▶ H_0 : person is infected; B : person is tested positive
- ▶ Known:
 - Sensitivity $P(B|H_0) = 95\%$ $P(\bar{B}|H_0) = 5\%$
 - Specificity $P(\bar{B}|\bar{H}_0) = 99\%$, $P(B|\bar{H}_0) = 1\%$
 - Incidence $P(H_0) = 5\%$

Bayes: pos. rate: $P(B) = P(B|H_0)P(H_0) + P(B|\bar{H}_0)P(\bar{H}_0) = 9.7\%$,

H_0 after pos. test: $P(H_0|B) = P(B|H_0)P(H_0) / P(B) = 51.0\%$

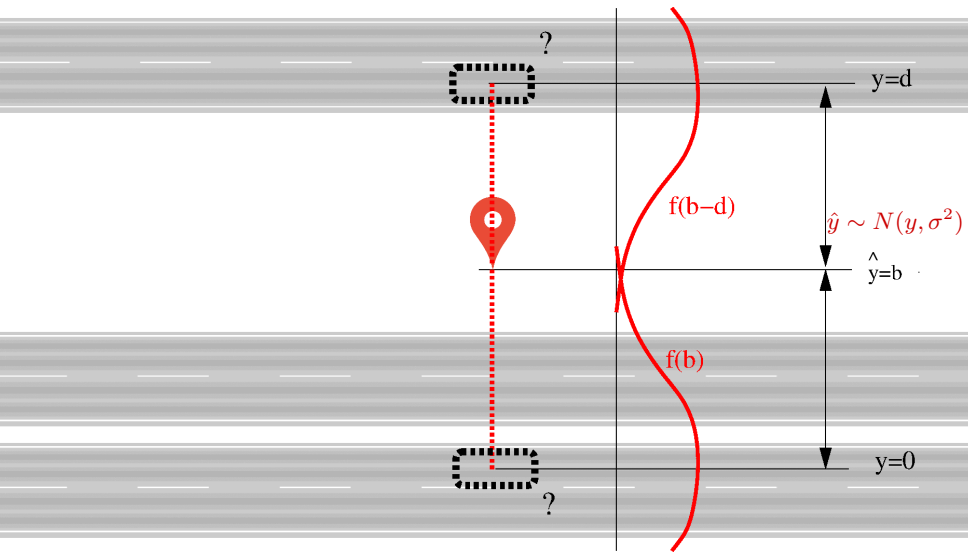
H_0 after neg. test: $P(H_0|\bar{B}) = P(\bar{B}|H_0)P(H_0) / P(\bar{B}) = 0.055\%$

5.4 Bayesian Inference for Discrete Quantities and Continuous Observations

- ▶ Discrete quantity/parameter β with the prior distribution
 $P(\beta = \beta_j) = p_j, \quad \sum_j p_j = 1$
- ▶ Continuous measurement $\hat{\beta}$ with a given distribution of density
 $g(\hat{\beta} \mid \beta = \beta_j) = f(\hat{\beta} - \beta_j)$
? What is the meaning of $f(\cdot)$? ! density of estimation error
- ▶ Assume $H_0 : \beta = \beta_{j_0}$ with $\beta_{j_0} \in \{\beta_j\}$ and the observation B :
 $\hat{\beta} \in [b - \delta/2, b + \delta/2]$ with arbitrarily small δ :
- ▶ Bayes: $P(H_0) = p_{j_0}$, $P(B|H_0) = \delta f(b - \beta_{j_0})$, and
 $P(B) = \delta \sum_j p_j f(b - \beta_j)$

$$\Rightarrow P(H_0 | \hat{\beta} = b) = \frac{P(H_0)P(B|H_0)}{P(B)} = \frac{p_{j_0} f(b - \beta_{j_0})}{\sum_j p_j f(b - \beta_j)}$$

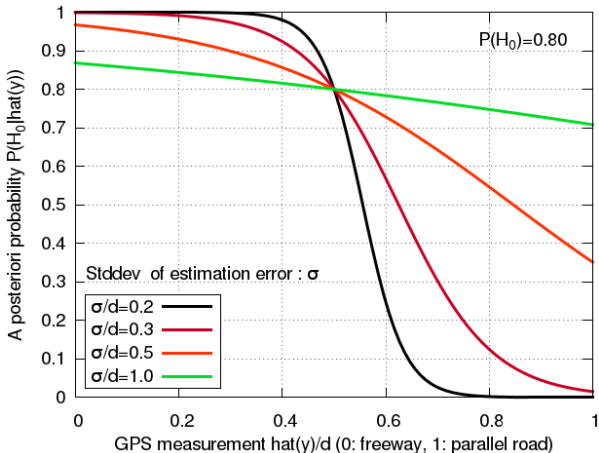
Example: Map matching



$$p(H_0) = \frac{\text{density freeway}}{\text{density freeway} + \text{density road}} = 0.8$$

$$P(H_0 | \hat{y} = b) = \frac{0.8f(b)}{0.8f(b) + 0.2f(b-d)}$$

Map matching II



True vehicle position:

$$y = \begin{cases} 0 & \text{freeway} \\ d & \text{parallel road} \end{cases}$$

Lateral GPS measurement:

$$\hat{y} \sim \begin{cases} N(0, \sigma^2) & \text{freeway} \\ N(d, \sigma^2) & \text{road} \end{cases}$$

Measured:

$$\hat{y} = 30 \text{ m}, \sigma = 10 \text{ m},$$

at a distance $d = 50 \text{ m}$

Read from graphics:

$$\frac{\sigma}{d} = 0.2, \frac{\hat{y}}{d} = 0.6$$

$$\Rightarrow P(H_0|\hat{y}) = 0.23$$

\Rightarrow you are on the parallel road
with a probability of 77 %

5.5 Bayesian Inference for Continuous Quantities and Measurements

- ▶ The quantity β has the a-priori distribution density $h(\beta)$
- ▶ Unlike discrete quantities/parameters, H_0 needs to be an interval instead of a point (why?) $\Rightarrow P(H_0)$ and $P(B|H_0)$ are integrals over the values of H_0
- ▶ Probability for H_0 based on measurements lying in the extreme region of a given measurement ($B = E_{\text{data}}$):

$$\begin{aligned}
 P(H_0|E_{\text{data}}) &= \frac{P(E_{\text{data}}|H_0)P(H_0)}{P(E_{\text{data}})} \\
 &\stackrel{P(H_0) \rightarrow \int_{\beta \in H_0} h(\beta) \, d(\beta)}{=} \frac{\int_{\beta \in H_0} P(E_{\text{data}}|\beta)h(\beta) \, d\beta}{\int_{\beta \in \mathbb{R}} P(E_{\text{data}}|\beta)h(\beta) \, d\beta} \\
 &\stackrel{\text{def power function } \pi}{=} \frac{\int_{\beta \in H_0} \pi(\beta)h(\beta) \, d\beta}{\int_{\beta \in \mathbb{R}} \pi(\beta)h(\beta) \, d\beta}
 \end{aligned}$$

Inference for a given measurement

Probability for H_0 based on a given realisation (measurement)

$\hat{\beta} \in B = [b - \delta/2, b + \delta/2]$ with arbitrarily small δ :

- ▶ β has the a-priori distribution density $h(\beta)$
- ▶ The estimation error $\hat{\beta} - \beta$ is independent from β (as in the OLS estimator under Gauß-Markow conditions), so $\hat{\beta}$ has the conditional density $g(b|\beta) = f(b - \beta)$

$$P(H_0|B) = \frac{P(B|H_0)P(H_0)}{P(B)}$$

$$P(H_0) \rightarrow \int_{\underline{}}^{\overline{}} h(\beta) \, d(\beta) \quad \frac{\int_{\beta \in H_0} \delta g(b|\beta) h(\beta) \, d\beta}{\int_{\beta \in \mathbb{R}} \delta g(b|\beta) h(\beta) \, d\beta}$$

$$\Rightarrow P(H_0|B) = \frac{\int_{\beta \in H_0} f(b - \beta) h(\beta) \, d\beta}{\int_{\beta \in \mathbb{R}} f(b - \beta) h(\beta) \, d\beta}$$

Notice that the denominator is just the convolution $[f * g]$ at $\hat{\beta} = b$

Example: Gaussian Prior Distribution and Observations

- ▶ Prior $\beta \sim N(0, \sigma_\beta^2)$ (expectation=0 is no restriction)
- ▶ Unbiased estimator $\hat{\beta} \sim N(\beta, \sigma_b^2)$
- ▶ Null hypothesis $H_0: \beta \leq \beta_0$
 $\Rightarrow f(b - \beta) = \phi\left(\frac{b - \beta}{\sigma_b}\right)$, $h(\beta) = \phi\left(\frac{\beta}{\sigma_\beta}\right)$, $\int_{H_0} d\beta = \int_{-\infty}^{\beta_0} d\beta$
- ▶ Bayesian inference for H_0 under the observation $\hat{\beta} = b$ (long calc.):

$$P(H_0|\hat{\beta}) = \Phi\left(\frac{\beta_0 - \mu}{\sigma}\right), \quad \mu = b \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_b^2}, \quad \sigma = \frac{\sigma_\beta \sigma_b}{\sqrt{\sigma_\beta^2 + \sigma_b^2}}$$

- ▶ When expressing the observation in terms of the p value, $b = \beta_0 + \sigma_b \Phi^{-1}(1 - p)$ and β_0 in terms of $P(H_0)$, $\beta_0 = \sigma_\beta \Phi^{-1}(P(H_0))$ (**derive!**), this result is valid for *any* simple intervall null hypothesis for a single parameter β , *any* a-priori expectation $E(\beta)$, and *any* H_0 boundary value β_0
- ▶ If $\sigma_b^2 \ll \sigma_\beta^2$ and H_0 is an interval, we have $P(H_0|\hat{\beta}) \rightarrow p$
 \Rightarrow "resurrection" of the p -value!

Questions

? Show that, if the variance of the prior distribution is much larger than that of the measurement, we have $P(H_0|\hat{\beta}) \rightarrow p$ and, if it is much smaller, we have $P(H_0|\hat{\beta}) \rightarrow P(H_0)$

! Answer to the first question, $\sigma_\beta \gg \sigma_b$:

we have

$$\mu = b \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_b^2} = b \frac{1}{1 + \frac{\sigma_b^2}{\sigma_\beta^2}} \rightarrow b,$$

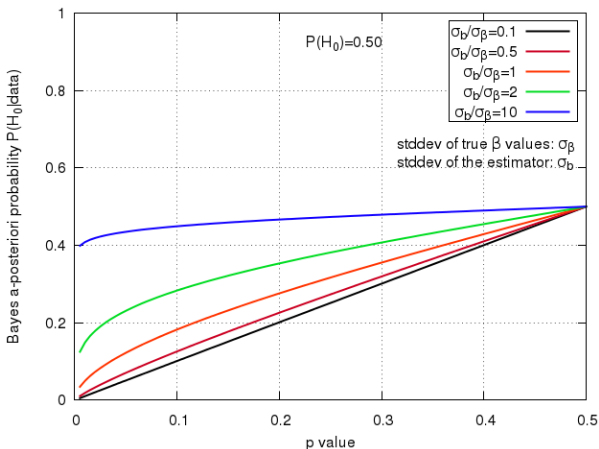
$$\sigma = \sigma_\beta \sigma_b / \sqrt{\sigma_\beta^2 + \sigma_b^2} = \sigma_b \sqrt{1 + \sigma_b^2 / \sigma_\beta^2} \rightarrow \sigma_b,$$

$$\begin{aligned} P(H_0|\hat{\beta}) &\rightarrow \Phi\left(\frac{\beta_0 - b}{\sigma_b}\right) \quad \beta_0 - b \stackrel{\text{in terms of } p}{=} \beta_0 - b \\ &= \Phi\left(-\Phi^{-1}(1-p)\right) \stackrel{\text{symm}}{=} \Phi\left(+\Phi^{-1}(p)\right) \stackrel{\text{def quantile}}{=} \underline{p} \quad \checkmark \end{aligned}$$

! Limiting case $\sigma_\beta \ll \sigma_b$: $\mu \rightarrow 0$, $\sigma \rightarrow \sigma_\beta$, $P(H_0|\hat{\beta}) = \Phi(\beta/\sigma_\beta) = P(H_0) \quad \checkmark$

Bayesian inference for a Gaussian prior distribution 1:

$P(H_0) = 0.5$



Example: Bike modal split β

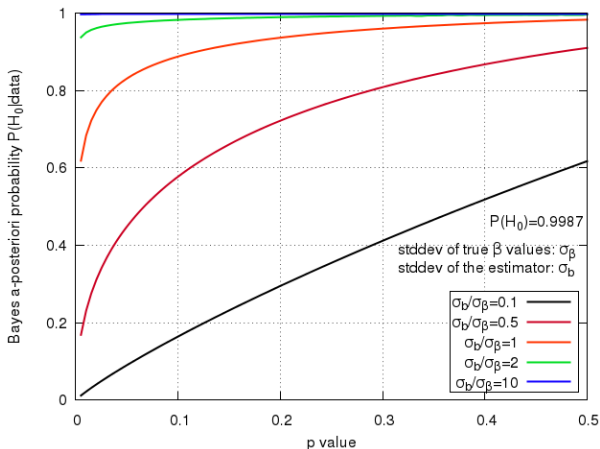
- ▶ Past investigation:
 $\beta = (20 \pm 3) \%$
- ▶ New investigation:
 $\hat{\beta} = (26 \pm 3) \%$

Has biking increased?

- ▶ Frequentist:
 $H_0 : \beta < 20 \%$,
 $p = \Phi(-2) = 0.0227$ ✓
- ▶ Bayesian:
 $\sigma_\beta = \sigma_b = 3 \%$,
 $p = 0.0227$, $P(H_0) = 0.5$
read from graphics:
 $P(H_0|\hat{\beta}) = 8\% \Rightarrow$ **no!**
(a difference test would give the same)

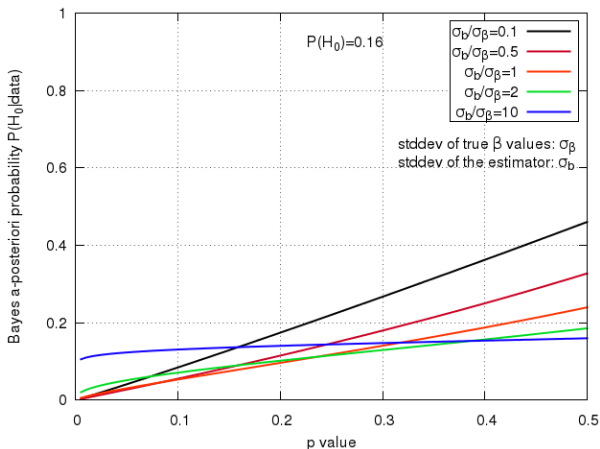
Bayesian inference for a Gaussian prior distribution 2:

$$P(H_0) = 0.9987$$



- ▶ $\sigma_b \ll \sigma_\beta$
 $\Rightarrow P(H_0|\hat{\beta}) \approx p$
 \Rightarrow precise a-posteriori information changes much.
- ▶ $\sigma_b \gg \sigma_\beta$
 $\Rightarrow P(H_0|\hat{\beta}) \approx P(H_0)$
 \Rightarrow fuzzy a-posteriori data essentially give no information \Rightarrow a-priori probability nearly unchanged.

Bayesian inference for a Gaussian prior distribution 3: $P(H_0) = 0.16$



Again, new data with $\sigma_b \ll \sigma_\beta$ gives much a-posteriori information (at least if p is significantly different from $P(H_0)$),

new data with $\sigma_b \gg \sigma_\beta$ are tantamount to essentially no new information.

5.6 Conclusion

- ▶ For discrete variables and measurements, we have the simple Bayes's calculations from elementary statistics → **probability tree**
- ▶ Discrete variables and continuous measurements:
 - If the measuring uncertainty is larger than the distance between possible discrete true values, then the **a-priori probability** matters
 - If the uncertainty is much smaller, then the **closest distance** to the measurement matters
 - The **p value is completely misleading**, even for bimodal continuous variables (vehicle not exactly in the middle of the right lane)
- ▶ Continuous variables and measurements:
 - The p value only gives a good estimate for the posterior probability $P(H_0|B)$ if (i) the prior distribution is **unimodal**, (ii) the measuring uncertainty is **much smaller** than the prior standard deviation, (iii) we have an **interval** null hypothesis
 - If the measuring uncertainty is much larger than the prior spread, the measurement **hardly changes** $P(H_0)$