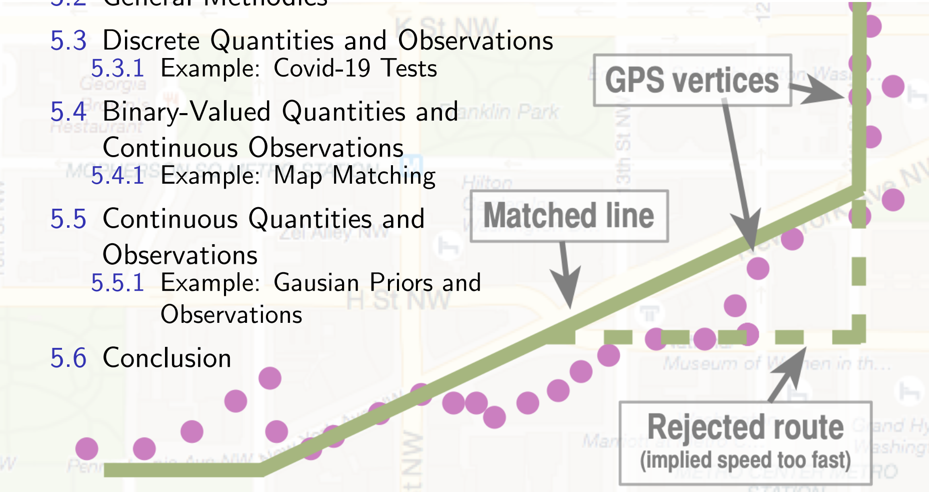
5. Is the p value dead? Frequentist vs. Bayesian inference



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 caculate 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 often 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 and continuous observation $\hat{\beta}$ (e.g., regression models)

5.3 Bayesian Inference for Discrete Quantities and Observations

Textbook case: binary variables $\in \{$"true", "false"$\}$ (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 $P(B|H_0) = 95\,\%$ $\quad P(\bar{B}|H_0) = 5\,\%$
  - Specificity $P(\bar{B}|\bar{H}_0) = 97\,\%$, $\quad P(B|\bar{H}_0) = 3\,\%$
  - Incidence $P(H_0) = 5\,\%$
- Bayes:
  - Test incidence: $P(B) = P(B|H_0)P(H_0) + P(B|\bar{H}_0)P(\bar{H}_0) = 7.6\,\%$
  - H_0 after test positive: $P(H_0|B) = P(B|H_0)P(H_0) \,/\, P(B) = 62.5\,\%$
  - H_0 after test negative: $P(H_0|\bar{B}) = P(\bar{B}|H_0)P(H_0) \,/\, P(\bar{B}) = 0.27\,\%$

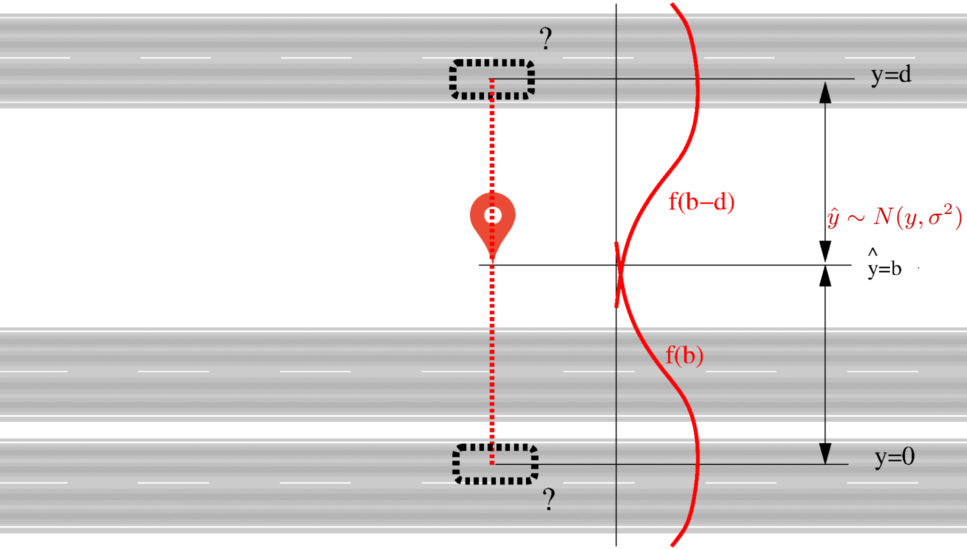
5.4 Bayesian Inference for Discrete Quantities and Continuous Observations

- 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 $f(.)$? ! 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 $P(B) = \delta \sum_j p_j f(b - \beta_j)$

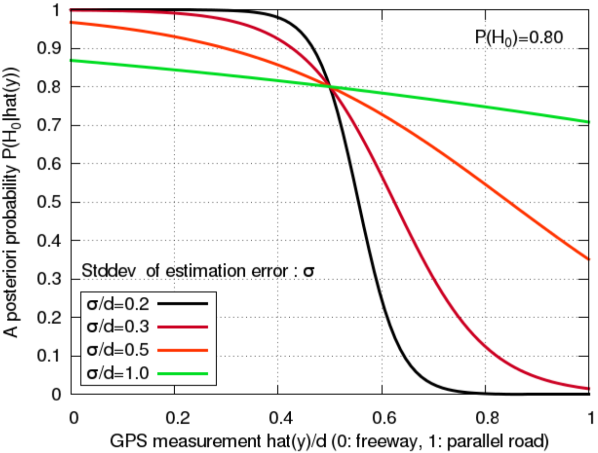
$$\Rightarrow \quad P(H_0|\hat{\beta} = b) = \frac{P(H_0)P(B|H_0)}{P(B)} = \frac{p_{j_0} f(b - \beta_{j_0})}{\sum_j p_j f(b - \beta_j)}$$

Example: Map matching



$$p(H_0) = \frac{\text{density freeway}}{\text{density freeway+density road}} = 0.8 \qquad P(H_0|\hat{y}=b) = \frac{0.8f(b)}{0.8f(b)+0.2f(b-d)}$$

Map matching II



True vehicle position:

$$y = \begin{cases} 0 & \text{freeway} \\ d = 50\,\mathrm{m} & \text{parallel road} \end{cases}$$

Lateral GPS measurement:

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

Measured:

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

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

Relation of Bayesian inference to the p-value and the power function

Probability for H_0 based on measurements lying in the extreme region of a given measurement ($B = E_{\text{data}}$):

$$P(H_0|E_{\text{data}}) = \frac{P(E_{\text{data}}|H_0)P(H_0)}{P(E_{\text{data}})} \overset{P(H_0)\to\int_{\beta\in H_0} h(\beta)\ \mathrm{d}(\beta)}{=} \frac{\int_{\beta\in H_0} P(E_{\text{data}}|\beta)h(\beta)\ \mathrm{d}\beta}{\int_{\beta\in \mathbb{R}} P(E_{\text{data}}|\beta)h(\beta)\ \mathrm{d}\beta}$$

$P(E_{\text{data}}|\beta)$ is related to the p-value $P(E_{\text{data}}|\beta_0 \in H_0^*)$ and also to the power function $\pi_\alpha(\beta) = P(R_\alpha|\beta)$ [$R_\alpha=$ rejection region at α]

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 $g(b|\beta) = f(b - \beta)$

$$P(H_0|B) = \frac{P(B|H_0)P(H_0)}{P(B)} \overset{P(H_0)\to\int h(\beta)\,\mathrm{d}(\beta)}{=} \frac{\int_{\beta\in H_0} \delta\ g(b|\beta)h(\beta)\ \mathrm{d}\beta}{\int_{\beta\in \mathbb{R}} \delta\ g(b|\beta)h(\beta)\ \mathrm{d}\beta}$$

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

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

Example: Gaussian Prior Distribution and Observations

- Prior $\beta \sim N(0, \sigma_\beta^2)$ (maximum ignorance), so $\beta/\sigma_\beta \sim N(0,1)$
- Unbiased estimator $\hat{\beta} \sim N(\beta, \sigma_b^2)$, so $(b-\beta)/\sigma_\beta \sim N(0,1)$
- Null hypothesis H_0: $\beta \le \beta_0$, so $\int_{H_0} \, \mathrm{d}\beta \, = \int_{-\infty}^{\beta_0} \, \mathrm{d}\beta$
- 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 $P(H_0|\hat{\beta}) \to p$
  $\Rightarrow$ "ressurrection" of the p-value!

Questions

? Show that, on the previous slide, $b = \beta_0 + \sigma_b \Phi^{-1}(1-p)$

! We assume known variance, so $T = (\hat{\beta} - \beta_0)/\sigma_b \sim N(0,1)$. For H_0: $\beta \leq \beta_0$ we have

$$\begin{aligned} p &= 1 - \Phi(t_{\text{data}}) \\ &= 1 - \Phi\left(\frac{b-\beta_0}{\sigma_b}\right) \\ \Phi\left(\frac{b-\beta_0}{\sigma_b}\right) &= 1-p \\ \frac{b-\beta_0}{\sigma_b} &= \Phi^{-1}(1-p) \\ b &= \beta_0 + \sigma_b \Phi^{-1}(1-p) \end{aligned}$$

? Show that, on the previous slide, $\beta_0 = \sigma_\beta \Phi^{-1}(P(H_0))$

! We have $P(H_0) = P(\beta \leq \beta_0) = \Phi\left(\frac{\beta_0}{\sigma_\beta}\right)$, so $\Phi^{-1}(P(H_0)) = \beta_0/\sigma_\beta$.

Questions II

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

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

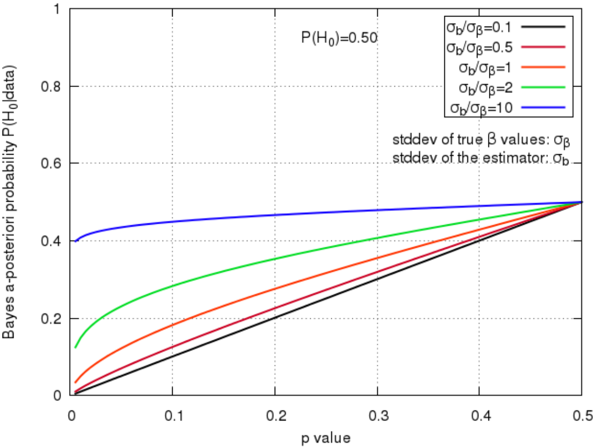
we have

$$\begin{aligned}
\mu &= b\frac{\sigma_\beta^2}{\sigma_\beta^2+\sigma_b^2} = b\frac{1}{1+\frac{\sigma_b^2}{\sigma_\beta^2}} \to b, \\
\sigma &= \sigma_\beta\sigma_b/\sqrt{\sigma_\beta^2+\sigma_b^2} \;=\; \sigma_b\sqrt{1+\sigma_b^2/\sigma_\beta^2} \to \sigma_b\,, \\
P(H_0|\hat{\beta}) &\to \Phi\left(\frac{\beta_0-b}{\sigma_b}\right) \overset{\beta_0-b \text{ in terms of } p}{=} \Phi\left(\frac{-\sigma_b\Phi^{-1}(1-p)}{\sigma_b}\right) \\
&= \Phi\left(-\Phi^{-1}(1-p)\right) \overset{\text{symm}}{=} \Phi\left(+\Phi^{-1}(p)\right) \overset{\text{def quantile}}{=} \underline{\underline{p}}
\end{aligned}$$

! Answer to the second question, $\sigma_\beta \ll \sigma_b$:

we have $\mu \to 0$, $\sigma \to \sigma_\beta$, $P(H_0|\hat{\beta}) = \Phi(\beta/\sigma_\beta) = P(H_0)$

Bayesian inference for a Gaussian prior distribution 1: $P(H_0) = 0.5$



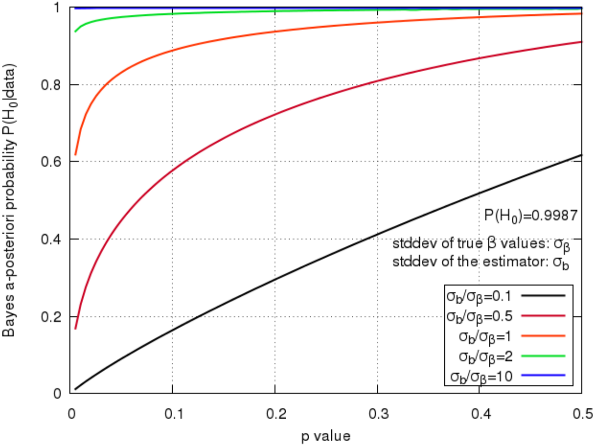
Example: Bike modal split β

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

Has biking increased?

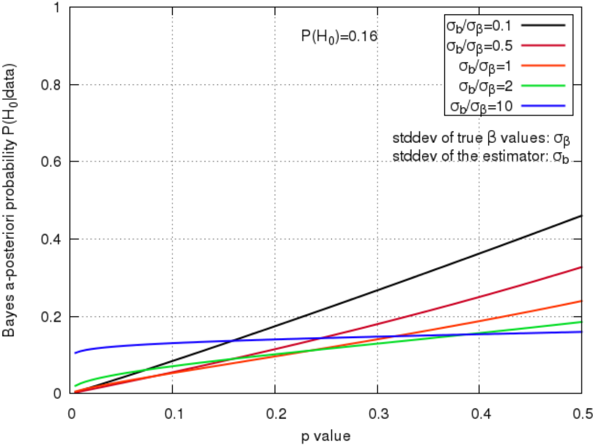
- Frequentist: $H_0 : \beta < 20\,\%$, $p = \Phi(-2) = 0.0227$ ✓
- Bayesian: $\sigma_\beta = \sigma_b = 3\,\%$, $p = 0.0227,\ P(H_0) = 0.5$ read from graphics: $P(H_0|\hat{\beta}) = 8\,\% \Rightarrow$ no! (a difference test would give the same)

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-posteri information changes much.
- $\sigma_b \gg \sigma_\beta$
  $\Rightarrow P(H_0|\hat{\beta}) \approx P(H_0)$
  $\Rightarrow$ fuzzy a-posteri 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 $\rightarrow$ 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 mislading, 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 $P(H_0)$