# Methods in Transportation Econometrics and Statistics (Master)

## Winter semester 2023/24, Solutions to Tutorial No. 5

### Solution to Problem 5.1: Map matching

(a) The only a-priori information is the traffic volume which can roughly translated to traffic density, i.e., the number of vehicles on a certain section. Obviously, if there are four times more vehicles on the freeway than on the parallel road, the probability for a random pick of any of these vehicles on the freeway is four times that on the parallel road as well. Hence

$$P(H_0) = 0.8, \quad P(H_1) = 0.2.$$

(b) There are further hints which, however, cannot qualify as a priori information. For example, a vehicle cannot *jump* between the roads. Furthermore, roads are only parallel to others for a finite distance. If the navigation device were switched on earlier, the GPS receiver would have picked up signals at locations where the two roads are further away from each other allowing a disambiguation.

(c) Classical frequentist statistics:

(i) $H_0$: $y = 0$

(ii) Known variance: $T = (\hat{y} - 0)/\sigma \sim N(0, 1)$

(iii) Data realisation: $t_{\text{data}} = 2$

(iv) Decision: Rejection range at $\alpha$: $R(\alpha) = \{t : t > t_{1-\alpha/2}\}$, so

$$\text{rejection if } t_{\text{data}} \in R(\alpha) \text{ or } 2 > t_{1-\alpha/2} = 1.96$$

(iva) Or calculation of $p$-value using the extreme range $E_{\text{data}} = \{t : t > 2\}$,

$$p = P(E_{\text{data}}|H_0^*) = P(E_{\text{data}}|H_0) = P(|T| > 2) = 2(1 - \Phi(2)) = 0.0454$$

Either way, $H_0$ is rejected at $\alpha = 5\,\%$, so frequentists would conclude $H_1$, i.e., the vehicle is on the parallel road

(d) Calculating the a-priori information necessary for the Bayes approach:

- Measuring event $M$: $\hat{y} \in [20 - \delta/2, 20 + \delta/2]$
- Probability of $M$ if $H_0$ making use of the smallness of the interval with $\delta$ and the Gaussian probability density $f_0(y)$ of the GPS value under $H_0$:

$$P(M|H_0) = \int\limits_{20-\delta/2}^{20+\delta/2} f_{\hat{y}}(y)\, \mathrm{d}y \approx \delta f_0(20)$$

With

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

and $\sigma = 10\,\mathrm{m}$, this results in

$$P(M|H_0) = 0.00540\ \delta$$

– The probability of $M$ under $H_1$ is calculated similarly:

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-50)^2}{2\sigma^2}\right), \\ P(M|H_1) &= \delta f_1(20) = 0.000443\ \delta \end{aligned}$$

and, with the definition of conditional probabilities and the addition of probabilities of disjunct events,

$$P(M) = P(H_0)P(M|H_0) + P(H_1)P(M|H_1) = 0.00441\ \delta$$

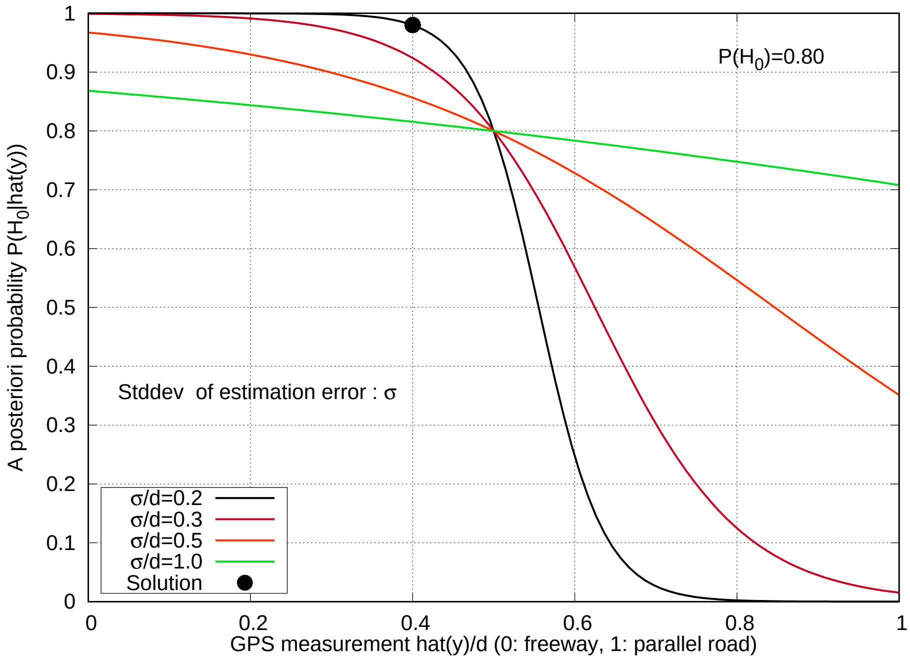
- Bayes Theorem:

$$P(H_0|M) = \frac{P(M|H_0)P(H_0)}{P(M)} = \frac{0.00540 * 0.8}{0.00441} = 0.980.$$

Notice that the auxilliary variable $\delta$ cancels out.

*Conclusion:* In reality, the car is on the freeway with a probability 98 % which is markedly different from the probability of less than 5 % what one could erroneously conclude from the $p$-value of the test for $H_0$

The following figure illustrates the Bayes probabbilities as a function of the relative position measurement value $\hat{y}/d$ and the relative uncertainty $\sigma/d$ for $P(H_0) = 0.8$. Here, we have $\hat{y}/d = 0.4$ and $\sigma/d = 0.2$.




## Solution to Problem 5.2: Frequentist vs. Bayes inference for Gaussian errors

(a)
1. $H_0 : f \leq 16\,\%$
2. $T = \frac{\hat{f}-16}{2} \sim N(0,1)$ if $H_0^* : f = 16\,\%$
3. $t_{\text{data}} = \frac{19-16}{2} = 1.5$
4. $E_{\text{data}} = \{t : t > 1.5\}$, $p = P(E_{\text{data}}|H_0^*) = P(T > 1.5) = 1 - \Phi(1.5) = 6.7\,\%$.

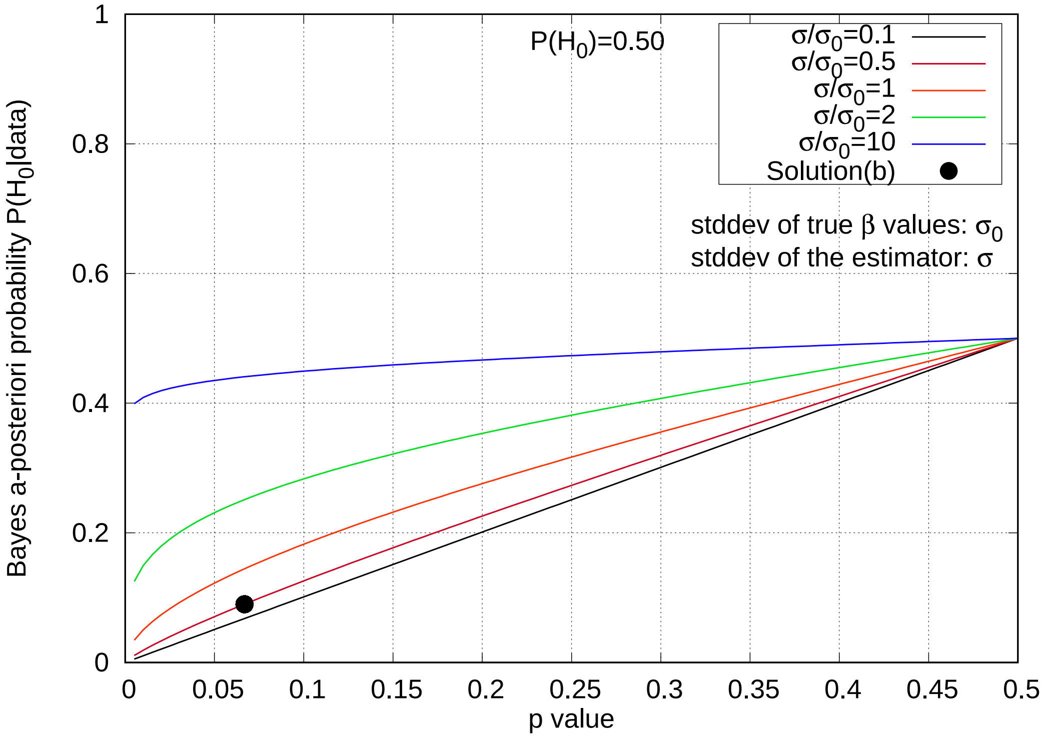
(b) The Prior distribution for the modal split, $f_0$, is given by $f_0 \sim N(16, 4^2)$, so

$$P(H_0) = P(f_0 \leq 16) = F_{f_0}(16) = \Phi(0) = 0.5.$$

The Bayes posterior can be obtained from the first graphics for $P(H_0)$. We have $\sigma/\sigma_0 = 0.5$, so the red curve is relevant. Reading it off at $p = 6.7\,\%$, we obtain

$$P(H_0|\hat{f}) \approx 10\,\%$$

The high modal split $\hat{f} = 19\,\%$ compared to the expected prior $\mu = 16\,\%$ decreases the prior probability for $H_0$: $f \leq 16\,\%$ from $50\,\%$ to $10\,\%$. However, this value is greater than the $p$ value, so the prior somewhat influences the result.

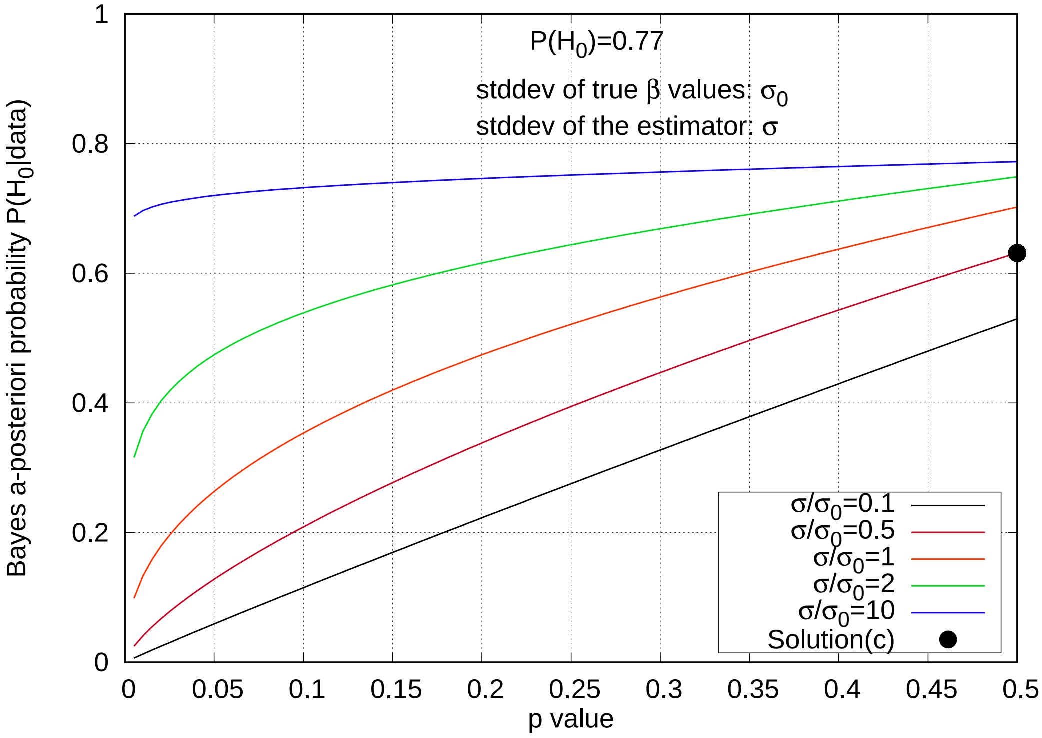



(c) The new null hypothesis is $H_{02} : f \le 19\,\%$

$$
\begin{aligned}
P(H_{02}) &= F_{f_0}(19) = \Phi\left(\frac{19-16}{4}\right) = 77\,\% \;\Rightarrow\; \text{use second diagram}\\
p &= 1-\Phi\left(\frac{19-19}{2}\right) = 1-\Phi(0) = 50\,\%\\
P(H_{02}|\hat{f}) &\approx 62\,\%
\end{aligned}
$$

Again, the comparatively high measured value decreases the probability for $H_{02}$ from its prior value 77 % to the posterior value 62 % which is, however, distinctly higher than the $p$ value 50 %.

Without a-priori information, i.e., $\sigma/\sigma_0 \to 0$, the posterior probability is equal to the $p$ value of 50 %. (The black curve for $\sigma/\sigma_0 \to 0.1$ yielding 52 % comes already close to that.)




## Solution to Problem 5.3: Mobility Survey

(a) Design:

- Aggregation level: disaggregated w/respect to age, income, gender; no single-person data, however.
- Time-subject space: the same persons ("the panel") surveyed repeatedly $\Rightarrow$ Panel design
- Type of question: revealed choice/revealed preferences (RP): The attractiveness of the different options will be calculated indirectly by the realized choices.
- Drawing method: layered/stratified sample according to the proportions of age, gender, city size in the population .
- Modality: written survey, combined with telephone and individual internet survey (one-shot individual password).

(b) Panel survey since, unlike the procedure in a trend design, the same persons are surveyed multiple times.

(c) Delimitation of the population:

- spatial: Germany

- temporal: Since 1995, open-end
- by subject: Only city inhabitants in cities $> 200\,000$ inhabitants; age 12+

The basis for drawing (register of the city inhabitants) is a superset since this register also includes the younger children.

(d) Stratified/layered sample; the strata are the age classes, gender, and number of inhabitants of each city. All stratified properties are "good" in the sense that they are (i) known in the population, (ii) arguably influence the activity/trip/mode choices.

(e) Surveyed properties:

- General socioeconomic variables (including the strata): age, genderAlter, #people in houshold, profession, income, city.
- Mobility related socioeconomic variables: ownership/availability of bikes and cars, possession of season tickets, distance to the next bus/tram stop.
- Activity related variables: #trips to work, home, and elsewhere; starting times of the trips.
- Characteristics (alternative-specific attributes): distance, travel times, and ad-hoc costs of the relevant trip types for all eligible modes/means of transport.