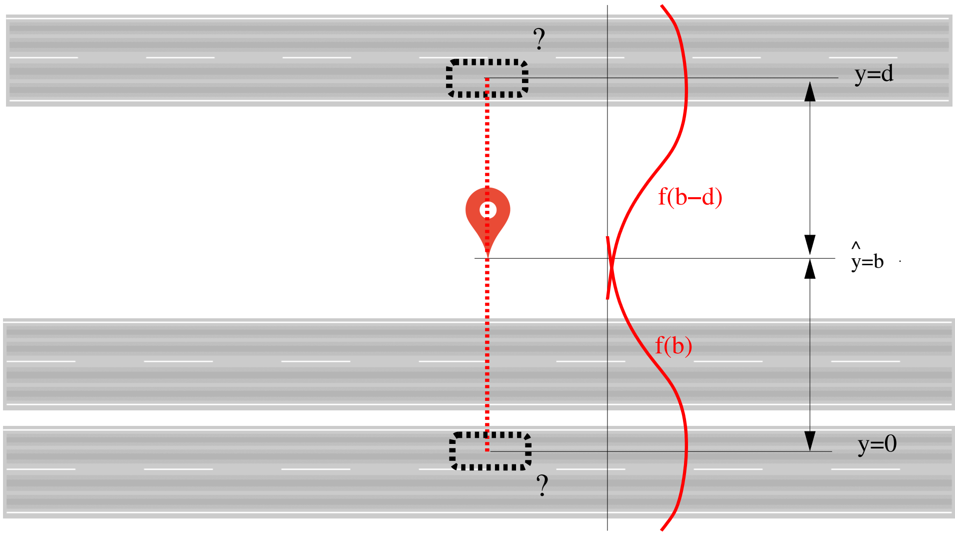
Methods in Transportation Econometrics and Statistics (Master)

Winter semester 2024/25, Tutorial No. 5

Problem 5.1: Map matching

In a classic map-matching problem, the navigation system matches the GPS position (isotropic positional errors of known variance $\epsilon \sim N(0, \sigma^2)$) to the nearest road by projecting it onto the axis of this road. However, if there are two parallel roads in a distance d, the result is ambiguous if d is not much greater than the positional error σ.



Consider the case of a freeway and a parallel road at a distance $d = 50$ m and assume an GPS error $\sigma = 10$ m. Define the coordinate y to be orthogonal to the two road axes with $y = 0$ at the freeway and $y = b$ at the other road. From past traffic measurements, it is known that, at the time of this trip, there is a traffic flow (traffic volume) of 4 000 veh./h on the freeway and 1 000 veh./h on the parallel road. The navigation device is just switched on.

(a) Give the a-priori probability distribution of the true vehicle position before the GPS receiver catches the first signal. Assume that the car is on the axis of one of the roads.

(b) Why would the a-priori distribution be different if the navigation device had been activated earlier?

(c) The GPS reading for the transversal coordinate is $\hat{y} = 20$ m. Calculate the p value for the null hypothesis H_0 “the vehicle drives on the freeway”. Could H_0 be rejected at an error probability $\alpha = 5\,\%$?

(d) Instead of the point measurement, define the mearuring event M: the GPS devise measures a transversal coordinate in the range $[\hat{y}-\delta/2, \hat{y}+\delta/2]$ with $\delta \ll \sigma$ such that a constant probability density can be used to determine probabilities. Calculate the conditional probabilties $P(M|H_0)$ and $P(M|H_1) = P(M|\overline{H_0})$ if the vehicle is on the freeway (H_0) or the parallel road (H_1), respectively. Also calculate the unconditional (a priori) probability for M.

Hint: the probability density $\phi(z) = 1/\sqrt{2\pi}\exp(-z^2/2)$ of the standard normal ditribution is related to the density $f(y)$ of $Y \sim N(\mu, \sigma^2)$ via

$$f(y) = \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right).$$

(e) Now use Bayes' Theorem to calculate the probability that the car is on the freeway or on the parallel road after the GPS measurement. To which road the navigation system should match the car?

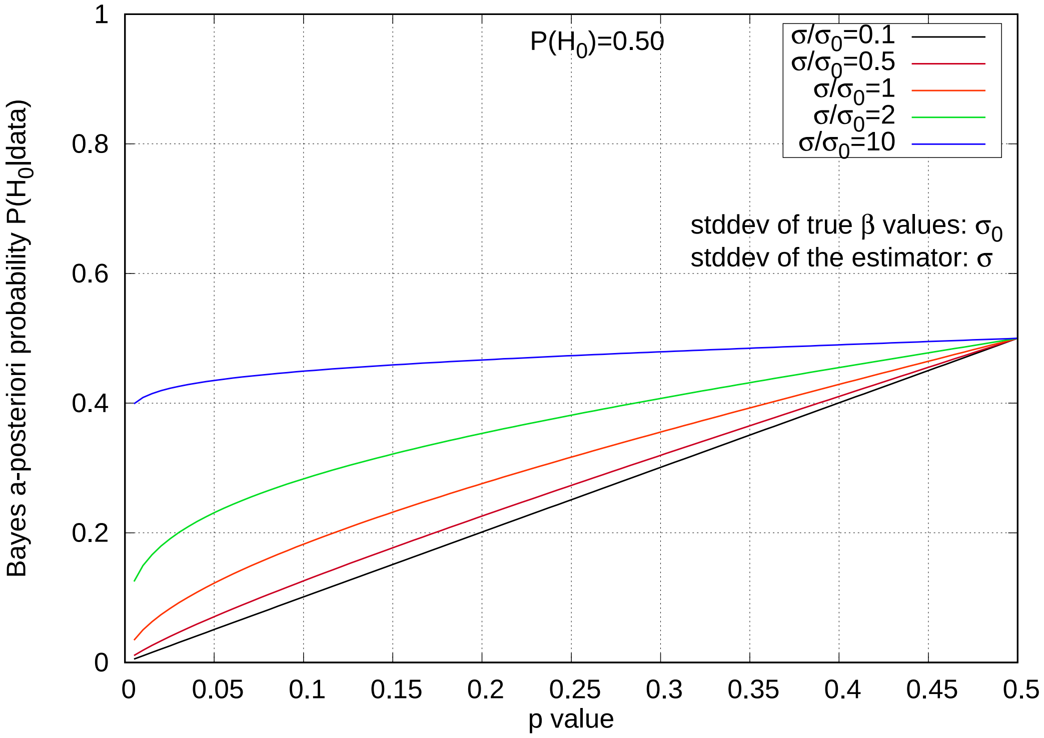
Problem 5.2: Frequentist vs. Bayes inference for Gaussian errors

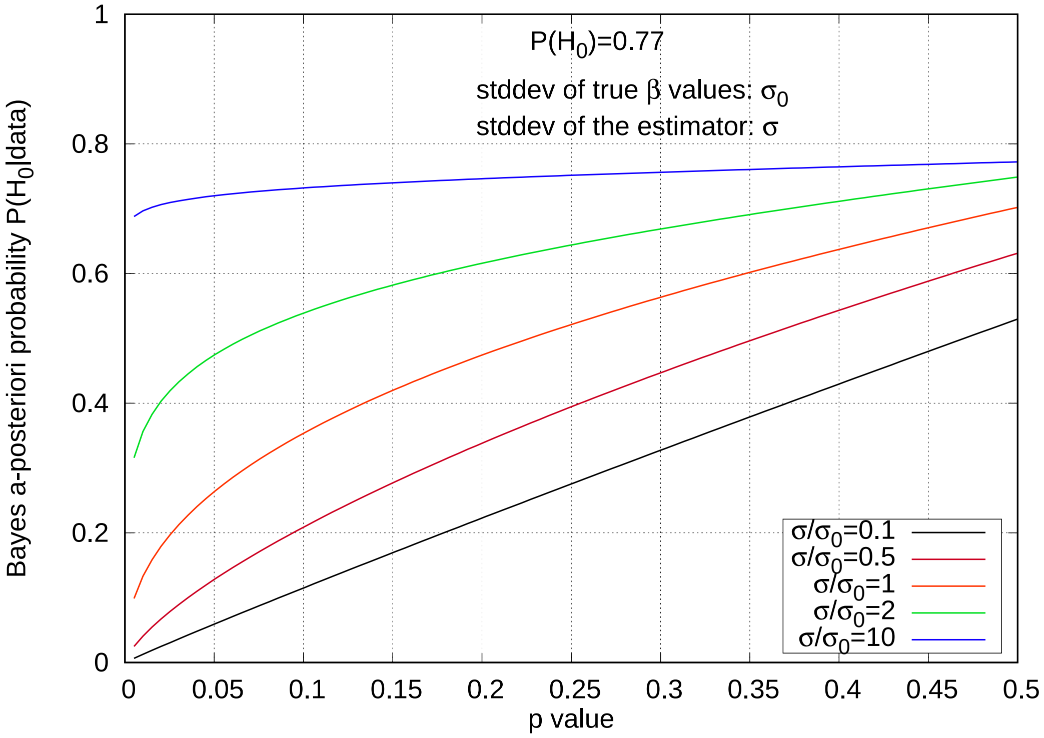
By means of a survey, a city administration estimates the fraction of cyclists in the modal split of this city to be $\hat{f} = 19\,\%$ at a (known) standard deviation $\sigma = 2\,\%$. In other comparable cities, the average bike fraction was $\mu_0 = 16\,\%$ and the variation from city to city (including true variations and estiating errors) was $\sigma_0 = 4\,\%$.

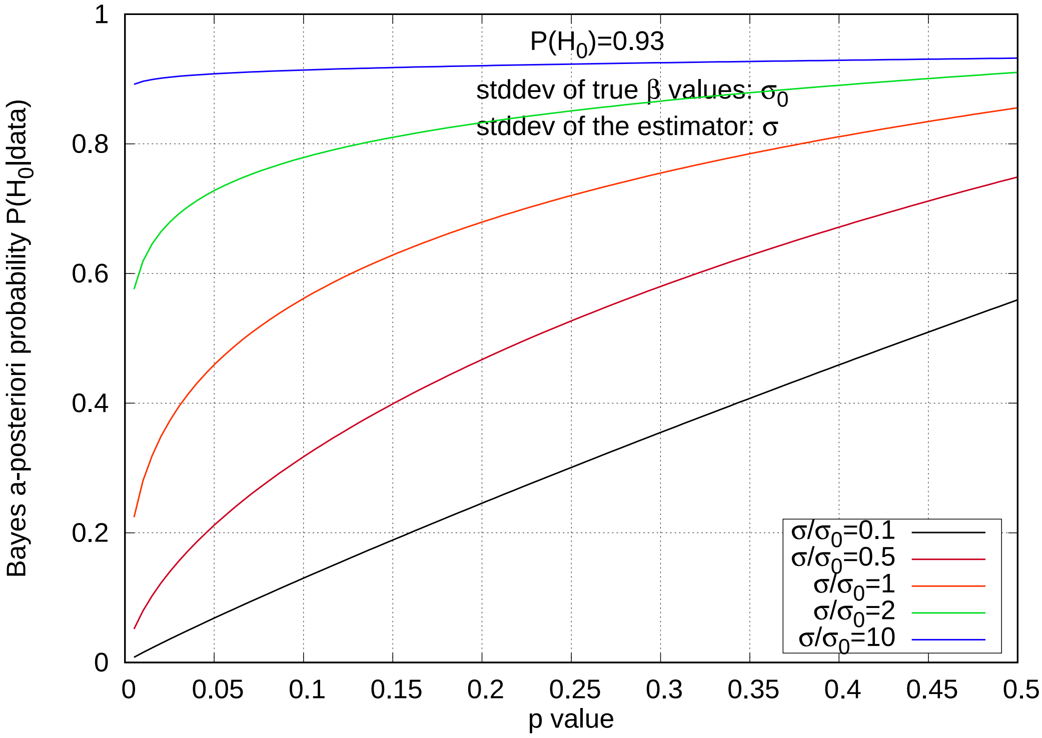
(a) Calculate the p value for the null hypothesis H_0: $f \le 16\,\%$

(b) Assume that, based on the modal splits in the other cities, the prior distribution f_0 of the bike percentage (before the survey) is $f_0 \sim N(\mu_0, \sigma_0^2)$. Show that the unconditional (prior) probability for the null hypothesis is given by $P(H_0) = 0.5$. Read off the Bayes posterior $P(H_0|\hat{f})$ from one of the following graphics.

(c) Determine, in the same way, the prior and posterior probabilities for the new null hypothesis $H_{02} : f \le 19\,\%$. Discuss the result. Now assume that not any prior observations or guesses are known. Does this change the probability?


P(H0)=0.50
σ/σ0=0.1
σ/σ0=0.5
σ/σ0=1
σ/σ0=2
σ/σ0=10
stddev of true β values: σ0
stddev of the estimator: σ
Bayes a-posteriori probability P(H0|data)
p value
0
0.2
0.4
0.6
0.8
1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5


P(H0)=0.77
stddev of true β values: σ0
stddev of the estimator: σ
σ/σ0=0.1
σ/σ0=0.5
σ/σ0=1
σ/σ0=2
σ/σ0=10
Bayes a-posteriori probability P(H0|data)
p value
0
0.2
0.4
0.6
0.8
1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5



Problem 5.3: Mobility Survey

By means of a long-term mobility survey one wants to find out the criteria influencing the atractiveness of the public transport in big German cities (200 000+ inhabitants). In particularly, long-term shift in the weighting of the criteria should be uncovered. Furthermore, the difference of these weightings for different socioeconomic groups are to be investigated.

In the beginning (1995) of the still active study, persons aged 12+ are drawn randomly from the register of the participating cities conditioned that the sample represents the gender and age disatribution of the population. Furthermore, the number of drawn persons is proportional to the city size. Later on, new participants were only acqired to replace drop-outs and to compensate for the shift in the age structure.

Each year, the surveyed participants filled out a written questionnaire. In case of no response, the persons are called and could answer the questions orally. Nowadays, also an individual electronic form protected by a one-off password is available.

The questionnair includes questions about the age, gender, #persons in the houshold, ownership/availability of bicycles and cars, possession of season tickets, location of home and work place, profession, income. Furthermore, the questionnaire includes questions on the favoured mode of transport for trips to work, to go shopping, and for other trips as well as the number of trips in each category and the typical start times.

In order to evaluate the attraciveness of different modes, the form contained questions about the distance to the next public transport stop, bus/tram/train headways at the starting time, duration of the respective trips with different modes, and the costs.

(a) classify the survey with respect to design, aggregation level, time-subject dimension, type of question, drawing method, and drawing modality.

(b) Both in the trend and panel designs, the questions are asked repeatedly in regular intervals. Why it is nevertheless clear that we deal with a panel design here?

(c) Delimit the population spatially, temporally, and by subject. Is the drawing basis identical with the population?

(d) Justify whether we have a random or stratified sample. If the latter, specify the strata.

(e) Name at least three properties/variables from each of the following categories
- General socioeconomic variables
- mobility-related socioeconomic variables,
- activity related variables,
- characteristics of the alternatives.