

Laboratory Session 06 : May 19, 2022
Exercises due : June 5, 2022

Exercise 1

- a well established and diffused method for detecting a disease in blood fails to detect the presence of disease in 15% of the patients that actually have the disease.
 - A young UniPD startUp has developed an innovative method of screening. During the qualification phase, a random sample of $n = 75$ patients known to have the disease is screened using the new method.
- (a) what is the probability distribution of y , the number of times the new method fails to detect the disease ?
- (b) on the $n = 75$ patients sample, the new method fails to detect the disease in $y = 6$ cases. What is the frequentist estimator of the failure probability of the new method ?
- (c) setup a bayesian computation of the posterior probability, assuming a beta distribution with mean value 0.15 and standard deviation 0.14. Plot the posterior distribution for y , and mark on the plot the mean value and variance
- (d) Perform a test of hypothesis assuming that if the probability of failing to detect the disease in ill patients is greater or equal than 15%, the new test is no better than the traditional method. Test the sample at a 5% level of significance in the Bayesian way.
- (e) Perform the same hypothesis test in the classical frequentist way.

Exercise 2

- Ladislaus Josephovich Bortkiewicz was a Russian economist and statistician. He noted that the Poisson distribution can be very useful in applied statistics when describing low-frequency events in a large population. In a famous example he showed that the number of deaths by horse kick among the Prussian army follows the Poisson distribution.
- Considering the following two sets of observations taken over a fixed large time interval in two different corps:

y death soldiers	0	1	2	3	4	≥ 5
n_1 observations	109	65	22	3	1	0
n_2 observations	144	91	32	11	2	0

- (a) assuming a uniform prior, compute and plot the posterior distribution for λ , the death rate over the measurement time. Determine the posterior mean, median and variance, and compute the 95% credibility interval.
- (b) assuming now a Jeffreys' prior,

$$g(\lambda) \propto 1/\sqrt{\lambda}, \text{ with } \lambda > 0$$

compute and plot the posterior distribution for λ , the death rate over the measurement time. Determine the posterior mean, median and variance, and compute the 95% credibility interval.

Exercise 3



- A study on water quality of streams, a high level of bacter X was defined as a level greater than 100 per 100 ml of stream water. $n = 116$ samples were taken from streams having a high environmental impact on pandas. Out of these, $y = 11$ had a high bacter X level.
- indicating with p the probability that a sample of water taken from the stream has a high bacter X level,

- (a) find the frequentist estimator for p
- (b) using a **Beta**(1, 10) prior for p , calculate and posterior distribution $P(p \mid y)$
- (c) find the bayesian estimator for p , the posterior mean and variance, and a 95% credible interval
- (d) test the hypotesis

$$H_o : p = 0.1 \text{ versus } H_1 : p \neq 0.1$$

at 5% level of significance with both the frequentist and bayesian approach

- a new measurement, performed one month later on $n = 165$ water samples, gives $y = 9$ high bacter X level
- (e) find the frequentist estimator for p
 - (f) find a bayesian estimator for p , assuming both a **Beta**(1, 10) prior for p , and assuming the posterior probability of the older measurement as the prior for the new one.
 - (g) find the bayesian estimator for p , the posterior mean and variance, and a 95% credible interval
 - (h) test the hypotesis

$$H_o : p = 0.1 \text{ versus } H_1 : p \neq 0.1$$

at 5% level of significance with both the frequentist and bayesian approach

Exercise 4

- analyze the data of Exercise 1 using a MCMC with JAGS (solve only point a of Ex 1)

Exercise 5

- analyze the data of Exercise 2 using a MCMC with JAGS

Exercise 6

- analyze the data of Exercise 3 using a MCMC with JAGS (solve point b and c)