EPFL, Spring 2023
Guillaume Obozinski

EE-209 Eléments de Statistiques pour les Data Sciences

Feuille d'exercices 6

Exercise 6.1 L'expérience sera répétée trois fois.

In various experimental sciences, it is common to repeat an experiment three times to get an idea of the uncertainty in the result. A chemist tries to determine the amount of a chemical element in water in millimoles per liter (mmol/l). He repeats the experiment which allows him to obtain a measurement three times and he obtains: 3.36, 2.67 and 2.65.

(a) Assuming that the data follows a Gaussian distribution with unknown mean and variance, determine a Student confidence interval at 95%. For convenience we provide some quantiles of Student distributions with different degrees of freedom. More precisely, if $t_\alpha^{(n)}$ denotes the quantile of the Student distribution with n degrees of freedom at level α, then we have:

$$t_{0.9}^{(1)} = 3.078, \quad t_{0.95}^{(1)} = 6.314, \quad t_{0.975}^{(1)} = 12.706,$$
$$t_{0.9}^{(2)} = 1.886, \quad t_{0.95}^{(2)} = 2.920, \quad t_{0.975}^{(2)} = 4.303,$$
$$t_{0.9}^{(3)} = 1.638, \quad t_{0.95}^{(3)} = 2.353, \quad t_{0.975}^{(3)} = 3.182\,.$$

Exercise 6.2 Luminosity of a star. We wish to estimate the luminosity of a distant star using a CCD sensor. The luminosity of the star can be related to the photon flux (average number of photons per second) θ hitting the detector given the distance of the star and the size of the detector. We will concentrate on estimating the photon flux θ. At each second we get a readout of the number of photons received from that star.

We obtain the following $n = 12$ observations: $62, 54, 56, 62, 62, 62, 54, 59, 51, 52, 84, 69$.

(a) What is the form of the maximum likelihood estimator for θ based on an i.i.d. sample of observations $x_1, \ldots, x_n$?

(b) Derive the expression of the Fisher information $I(\theta)$.

(c) Deduce from the previous question the form of a Wald confidence interval for θ.

(d) Calculate a 99% Wald confidence interval based on the data.

(e) Is correct to say that for the confidence interval that we have computed there is 99% chance that the true mean is within the interval?

Exercise 6.3 Time to breakdown. A theoretical model suggests that the time to breakdown of an insulating fluid between electrodes at a particular voltage has an exponential distribution with parameter λ. A random sample of $n = 50$ breakdown times has been obtained. The empirical mean of the sample is $\overline{x} = 42.8$ and the sample variance (unbiased variance estimate) is $s^2 = 2043.6$. We wish to obtain a 95% confidence interval for λ and for the average breakdown time $\mu = 1/\lambda$.

(a) Express the variance of an exponential random variable X as a function of λ.

(b) Write the form of the likelihood, the log-likelihood, and the score function in the exponential model for a sample of n i.i.d. observations.

(c) Determine the form of the maximum likelihood estimator for λ in the exponential model, based on n i.i.d. observations.

(d) Determine the form of the Fisher information $I(\lambda)$.

(e) Using the results from the previous questions, determine the form of a Wald confidence interval for λ.

(f) We now wish to estimate the average breakdown time $\mu = \frac{1}{\lambda}$; what is the maximum likelihood estimator for μ?

(g) Propose two different asymptotic confidence intervals for μ based on $\widehat{\mu}_{\text{MLE}}$, using, in one case, the classical CLT and, in the other, the CLT with Slutsky.

(h) Calculate the numerical value of these two confidence intervals for a nominal level of 0.95.

(i) ($**$) Compare these two asymptotic confidence intervals. Can you think of reasons to use one rather than the other?

(j) Use the two confidence intervals obtained in (h) to propose confidence intervals for λ.

(k) ($**$) Which confidence interval do you propose to use for λ in the end?