University of Lille 2023-2024
M1 Data Science - Stats 2 April 2nd, 2024

TD4 - (Likelihood Ratio) Testing

Exercise 1 We collect one sample X from a Poisson distribution with parameter λ. We recall that its probability mass function is given by

$$\forall k \in \mathbb{N}, \quad f_\lambda(x) = \frac{\lambda^k}{k!} e^{-\lambda} \; .$$

We want to test $\mathcal{H}_0 : (\lambda = 5)$ against $\mathcal{H}_1 : (\lambda = 10)$ at level $\alpha = 0.05$, based on X.

1. Prove that a randomized Neyman-Pearson test can be formulated as

$$\begin{aligned} \widetilde{D}(X) &= 1, \quad \text{if } X > t \\ \widetilde{D}(X) &= \gamma, \quad \text{if } X = t \\ \widetilde{D}(X) &= 0, \quad \text{if } X < t. \end{aligned}$$

2. Using that $P_{Z \sim \mathcal{P}(5)}(Z > 9) = 0.032$ and $P_{Z \sim \mathbb{P}(5)}(Z > 8) = 0.068$, deduce that $t = 9$ and $\gamma = 1/2$.

3. What is the power of this test?

Exercise 2 We collect iid data $X_1, \dots, X_n$ from an exponential distribution with parameter θ. We recall that its density is given by

$$\forall x \in \mathbb{R}, \quad f_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{[0,+\infty[}(x).$$

1. Propose a Uniformly More Powerful test of level α for the test

$$\mathcal{H}_0 : (\theta \le \theta_0) \quad \text{against} \quad \mathcal{H}_1 : (\theta > \theta_0) \; .$$

2. Can we propose a UMP(α) test for

$$\mathcal{H}_0 : (\theta = \theta_0) \quad \text{against} \quad \mathcal{H}_1 : (\theta \neq \theta_0) \; ?$$

Exercise 3 We consider a two-sample testing problem in which we observe $X_1, \dots, X_{n_1} \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y_1, \dots, Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma^2)$ where $(\mu_1, \mu_2) \in \mathbb{R}^2$ and want to test

$$\mathcal{H}_0 : (\mu_1 = \mu_2) \quad \text{against} \quad \mathcal{H}_1 : (\mu_1 \neq \mu_2)$$

We denote by $Z = (X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})$ the observation.

1. Assuming the variance σ^2 is known, compute

$$\log \widetilde{\Lambda}(Z) = \log \frac{\sup_{\mu_1, \mu_2} L(Z; \mu_1, \mu_2)}{\sup_{\mu_1, \mu_2 : \mu_1 = \mu_2} L(Z; \mu_1, \mu_2)} \; .$$

2. Propose a (non-asymptotic) LRT test of level α. Compute its power function.

3. Compare it with the asymptotic LRT test provided by Wilk's theorem.

Exercise 4 We consider a so-called one-parameter canonical exponential family, in which the density wrt to some reference measure is

$$f_\theta(x) = h(x)\exp(\theta x - b(\theta))$$

where b is some twice diffenrentiable function that is furhtermore stricly convex ($b'' > 0$). We admit that these assumption are sufficient to be in a regular model. We denote by $\mu(\theta) = \mathbb{E}_\theta[X]$ the expectation of the distribution parameterized by θ.

1. Prove that $b'(\theta) = \mathbb{E}_\theta[X]$ and $b''(\theta) = \mathrm{Var}_\theta[X]$.
2. Deduce that the mapping $\theta \mapsto \mu(\theta)$ is one-to-one. We denote by μ^{-1} its inverse.
3. Compute $\widehat{\theta}_n$, the maximum likelihood estimator of θ.
4. We introduce $\mathrm{K}(\theta, \theta')$, the Kullback-Leibler divergence between P_θ and $P_{\theta'}$, defined as

   $$\mathrm{K}(\theta, \theta') = \mathbb{E}_\theta\left[\log \frac{f_\theta(X)}{f_{\theta'}(X)}\right]$$

   Prove that

   $$K(\theta, \theta') = (\theta - \theta')\mu(\theta) - b(\theta) + b(\theta')$$

5. Deduce the following inequality: for all $\theta \in \Theta$,

   $$\log \frac{L(X_1, \dots, X_n; \widehat{\theta}_n)}{L(X_1, \dots, X_n; \theta)} = n\mathrm{K}(\widehat{\theta}_n, \theta)$$

6. Prove that the (generalized) log-likelihood ratio associated to the test

   $$\mathcal{H}_0 : (\theta \le \theta_0) \quad \text{against} \quad \mathcal{H}_1 : (\theta > \theta_0)$$

   satisfies

   $$\log \widetilde{\Lambda}(X) = n\mathrm{K}(\widehat{\theta}_n, \theta_0)\mathbb{1}(\widehat{\theta}_n \ge \theta_0).$$

7. We admit the following concentration inequality (called a Chernoff inequality):

   $$\forall \theta \in \Theta, \forall x > 0, \quad \mathbb{P}_\theta\left(\widehat{\theta}_n > \theta, n\mathrm{K}(\widehat{\theta}_n, \theta) > x\right) \le e^{-x} .$$

   Propose a LRT of level α. Is this test UMP(α)?

8. Propose a test of level α for testing

   $$\mathcal{H}_0 : (\mu = \mu_0) \quad \text{against} \quad \mathcal{H}_1 : (\mu \neq \mu_0) .$$

   Compare it to the asymptotic test of level α obtained using Wilk's theorem, for $\alpha = 0.05$. Which one will have the largest power?

Exercise 5 The scientist Mendel (considered as the father of genetics) did the following experiment. He bred two different kind of peas: one with round yellow seeds and one with wrinkled green seeds. There are four types of progeny: round yellow (1), wrinkled yellow (2), round green (3) and wrinkled green (4). For each individual in the progeny, we denote by p_i the probability that it is of type i.

Assuming that the individual are independent when there are n individuals, the number of individual of each kind $N_n = (N_{n,1}, N_{n,2}, N_{n,3}, N_{n,4})$ follows a so-called multinomial distribution with parameter n and $p = (p_1, p_2, p_3, p_4)$, for which

$$\mathbb{P}(N_n = (n_1, n_2, n_3, n_4)) = \frac{n!}{n_1! n_2! n_3! n_4!} \prod_{i=1}^{4} p_i^{n_i} .$$

Mendel's ineheritance theory predicts that p is equal to

$$p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right)$$

He did $n = 556$ experiments and observed $N_n = (315, 101, 108, 32)$. We want to test

$$\mathcal{H}_0 : (p = p_0) \text{ against } \mathcal{H}_1 : (p \neq p_0)$$

1. Perform a Likelihood Ratio Test of this hypothesis. Does it reject $\mathcal{H}_0$?

2. For the above testing problem with multinomial data it is common to use another test, called Pearson's χ^2 test. When there are k possible types, this test is based on the test statistic

$$T_n = \sum_{i=1}^{k} \frac{(N_{n,i} - np_{0,i})^2}{np_{0,i}}$$

which is proved to satisfy $T_n \rightsquigarrow \chi^2_{k-1}$. Perform a χ^2 test. Does it reject the hypothesis?

3. Do you think that using statistical testing is appropriate to valide a theory?

Exercise 6 Let Z denote a random variable with density

$$x \mapsto \frac{1}{\lambda} \exp\left(-\frac{x - \theta}{\lambda} \right) \cdot \mathbb{1}_{[\theta, +\infty[}(x),$$

where $\lambda > 0$ and $\theta \in \mathbb{R}$ are unknown. Let $(X_1, \ldots, X_n)$ be a n-sample of i.i.d. variables with the same distribution as that of Z.

1. What is the MLE $(\widehat{\lambda_n}, \widehat{\theta_n})$ of the unknown parameter (λ, θ)? Evaluate the bias of this estimator and provide an *unbiased* estimator of (λ, θ).

2. Assume that one is to test $\mathcal{H}_0 : (\lambda = 1)$ against $\mathcal{H}_1 : (\lambda \neq 1)$. Prove that the (Generalized) Likelihood Ratio Test rejects the null hypothesis $\lambda = 1$ whenever $\widehat{\lambda_n} \notin [a, b]$, where a and b satisfy some equations to be clarified.

3. Let us now assume that λ is known, and that $\alpha \in]0, 1[$ is given. Use the MLE of θ to build a confidence interval with confidence level $1 - \alpha$ for the parameter θ. Deduce a statistical test of $\mathcal{H}_0 :\ (\theta = \theta_0)$ against $\mathcal{H}_1 :\ (\theta \neq \theta_0)$.

4. Let us finally assume that θ is known and $\widehat{\lambda_n}$ denotes the MLE of λ. By using the fact that the exponential distribution with parameter $1/2$, i.e., $\mathcal{E}(1/2)$, is a χ^2 distribution with 2 degrees of freedom, what is the distribution of $2n\widehat{\lambda_n}/\lambda$? Deduce a confidence interval for λ with confidence level $1 - \alpha = 0.95$ when $n = 13$.