Assignment sheet 6

October 24, 2018

Assignment 1. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the $N(0, \sigma^2)$ distribution. Consider testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 = \sigma_1^2$, where $\sigma_0^2 > \sigma_1^2 > 0$ are fixed numbers.
(a) For a given significance level α, find the most powerful test using the Neyman–Pearson lemma. Simplify its critical region as much as possible and determine the critical value. Does the critical value depend on σ_1^2 ?
(b) For given σ_0^2, σ_1^2, n, α, calculate the power (rejection probability when H_1 holds) of the test. Discuss how it depends on σ_0^2, σ_1^2, n, α.
(c) For given σ_0^2, σ_1^2, α, determine the smallest number of observations needed to have power at least β (i.e., to reject H_0 with probability at least β when H_1 holds).

Assignment 2. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the $Ber(p)$ distribution. Consider testing $H_0 : p = p_0$ versus $H_1 : p = p_1$, where $0 < p_0 < p_1 < 1$ are fixed numbers.
(a) Using the Neyman–Pearson lemma, find for which values of α will there exist a most powerful test with significance level α. Simplify the critical region of any such most powerful test as much as possible and determine the critical value. Does the critical value depend on p_1 ?
(b) Suppose $p_0 = \frac{3}{10}$, $p_1 = \frac{1}{2}$, $n = 3$ and $\alpha = 0.05$. Does there exist a most powerful test at significance level α in this case ?
(c) Answer part (b) for $\alpha = 0.027$.
(d) Use the CLT to find a critical value such that the significance level of the test is approximately (asymptotically for large n) α.

Assignment 3. Let $X_1, \ldots, X_n \sim Exp(\lambda)$ independent. Suppose that we wish to test the *one-sided* hypothesis $H_0 : \lambda \leq 4$ versus $H_1 : \lambda > 4$. We shall see how to derive such a test based on the Neyman–Pearson lemma.
(a) Let $\lambda_1 = 5 > \lambda_0$. Find the form (that is, up to constant) of the optimal test for $H_0 : \lambda = 4$ versus $H_1 : \lambda = 5$. Try to simplify and use a test statistic, $T = T(X_1, \ldots, X_n)$ that is as simple as possible.
(b) Show that $\sum X_i \sim Gamma(n, \lambda)$. *Hint : use the moment generating function.*
(c) How would the test look like if we chose $\lambda_1 = 6$ instead of 5 ?
(d) Give a formula for the critical value of the test in terms of quantiles of a distribution you have seen in the course.
(e) We will now check the type one error of this test by means of simulations. Pick your favourite choice for $n \geq 2$ and generate a sample of size n from $Exp(4)$ distribution. Store this in a vector `X`.
(f) Use R to determine whether the test rejects $H_0 : \lambda \leq 4$ at significance level $\alpha = 0.05$. The R function `qgamma` can prove useful. Be CAREFUL to use the argument `rate` and not `scale` in `qgamma` in order to be consistent with the notation of the course.
(g) Repeat this experiment 1000 times. How many times do you expect H_0 to be rejected ? Verify your guess using R.
(h) Redo steps (e)–(g) where X is exponential but with parameter different than 4 (but the test still tests $H_0 : \lambda = 4$). What happens when the true parameter is 3 ? And when it is 5 ?

Assignment 4. A manufacturing company just purchased a new machine. A quality controller wishes to compare the number of fault pieces made per week by the new machine against the number they used to have. 100 independent observation across different locations yield a measurement of $\overline{x} = 20$ for the new machine against $\overline{y} = 22$ for the old production process.

Assume that the number of fault pieces per week has a Poisson distribution with mean θ_1 for the new machine and θ_2 for the old one.

(i). Write the likelihood ratio to test $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$ at a significance level $\alpha = 0.05$.

(ii). Compute the exact value of your likelihood ratio.

(iii). Use Wilks' theorem to find the approximate distribution of Λ. Based on this, find a critical value of Λ that would lead you to reject H_0.
(Read again slide 201).

Assignment 5. Let $X_1, \ldots, X_n \sim f(x;\theta)$, where f is 1-parameter exponential family. Another idea for building tests for bilateral hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_1$ is to directly compare θ_0 to the maximum likelihood estimator $\widehat{\theta}$, which we know enjoys good properties. We cannot simply use $(\widehat{\theta} - \theta_0)^2$, since such measure does not take into account the variability of $\widehat{\theta}$. We would like to use a standardised version $(\widehat{\theta} - \theta_0)^2 / var\widehat{\theta}$. Alas, the variance will typically depend on the unknown value of θ, and needs to be estimated as well. This results in what is known as a *Wald test.*
(a) What is the asymptotic variance of $\widehat{\theta}$? Call it $v(\theta)$ (recall that it depends on θ!)
(b) Since $\widehat{\theta}$ is close to θ, it makes sense to estimate $v(\theta)$ by $v(\widehat{\theta})$. We then obtain the *Wald test statistic*

$$T = \frac{(\widehat{\theta} - \theta_0)^2}{v(\widehat{\theta})}$$

Write a formula for T by plugging in the function v.
(c) Assuming that v is continuous and $v(\theta) > 0$, find the asymptotic distribution of T. *Hint : $v(\widehat{\theta})/v(\theta) \to 1$ in probability ; use Slutsky's theorem.*
(d) Let $X_1, \ldots, X_n$ be independent $N(0, \sigma^2)$ random variables and let $\sigma_0^2 > 0$. Use an approximate Wald test to test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$ at significance level α.
(e) What is the likelihood ratio test in this case ? Is it the same ?

Assignment 6. A car manufacturing company wishes to publish information on the oil usage of a new car model. For 12 cars of this new model they measure the liters necessary to run for a 100 km, with the following results

$$14.60, 11.21, 11.56, 11.37, 13.68, 15.07, 11.06, 16.58, 13.37, 15.98, 12.07, 13.22.$$

The empirical mean and variance of the sample are $\bar{x}_{12} = 13.31$ and $s_{12}^2 = \frac{1}{11}\sum_{i=1}^{12}(x_i - \bar{x})^2 = 3.69$, respectively. Suppose the sample is normally distributed. We want to test wether the mean consumption is equal to 12.2 litres vs the alternative hypothesis that the mean is different from 12.2 litres.

(a) Write the model, the null and the alternative hypothesis.
(b) Which test statistics would you use ?
(c) Which values of the test statistics would be considered "extremes" ?
(d) Use the test statistics to test at a 5 % significance level.
(e) Repeat for a 10 % significance level. Did you find any difference ?
(f) Compute the p-value p_{obs}.
(g) Repeat (**??**) et (**??**) using an approach based on the p-value p_{obs}.