Assignment sheet 5

October 18, 2017

Assignment 1. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the $N(\mu, 1)$ distribution. Let $\widehat{\mu}$ be the MLE of μ.
(a) Find $\widehat{\mu}$.
(b) Find the asymptotic distribution of $\widehat{\mu}$.
(c) Using part (b) and without direct calculations, find the Cramer-Rao lower bound for the variance of an unbiased estimator of μ.
(d) Is there an estimator that satisfies this lower bound for each fixed n?
(e) Suppose that we are interested in estimating $g(\mu) = \mathbb{P}[X_1 \leq 2]$. Find an explicit expression for $g(\mu)$.
(f) Find the MLE of $g(\mu)$. Denote it by T.
(g) Using the delta method, find the asymptotic distribution of T.

Assignment 2. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the distribution with density function

$$f_X(x) = \begin{cases} \frac{\alpha\pi^\alpha}{x^{\alpha+1}}, & x \geq \pi \\ 0 & x < \pi. \end{cases}$$

(This is a Pareto distribution and α is called the tail index or the Pareto index.)
(a) Find the MLE $\widehat{\alpha}$ of α.
(b) Find the asymptotic distribution of $\widehat{\alpha}$.
(c) Let $Y = \log(X/\pi)$. Find the distribution of Y directly, i.e., without using transformation of variables.
(d) Find the asymptotic distribution of $T(Y_1, Y_2, \ldots, Y_n) := \sum_{i=1}^n Y_i$.
(e) Express $\widehat{\alpha}$ in terms of $T(Y_1, Y_2, \ldots, Y_n)$, and use this along with part (d) to find the asymptotic distribution of $\widehat{\alpha}$.
(*Hint : Use the delta method.*)

Assignment 3. 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 4. 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 5. 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 6. In this assignment we shall see empirically that Stein's estimator has a lower mean squared error than the maximum likelihood estimator.
Let y_1, y_2 and y_3 be independent normal random variables with unit variance and unknown means μ_1, μ_2 and μ_3.
(a) Use R to simulate one realisation of the random vector $y = (y_1, y_2, y_3)$ for the parameter value $\mu = (\mu_1, \mu_2, \mu_3) = (-1, 0, 1)$. *Hint : the command* `rnorm` *can take vector values.*
(b) What is the optimal value of a in terms of the mean squared error of the James–Stein estimator $\widetilde{\mu}_a$? Write an R command that calculates it, for a sample stored in a vector $Y \in \mathbb{R}^3$. *Hint : you can use* `sum(Y^2)`.
(c) Repeat the simulation 1000 times. For each repetition, calculate the errors $\|\mu - \widehat{\mu}\|^2$ and $\|\mu - \widetilde{\mu}_a\|^2$. Store these in two vectors of length 1000, `MSE.mle` and `MSE.stein`. Use these vectors to approximate the mean squared error of the two estimators. Which one is smaller ? Try changing the values of μ (and perhaps n and a).

Assignment 7. In this assignment we give an alternative approach to shrinkage by means of adding a penalty term to the optimisation problem.
(a*) Let $X \sim Gamma(k, \lambda)$ with $k > 1$. Using the property that $\Gamma(x) = (x-1)\Gamma(x-1)$ for $x > 1$, show that $\mathbb{E}\,\frac{1}{X} = \lambda/(k-1)$.
(b) Using part (a), show that if $X \sim \chi^2_n$ with $n > 2$, then $\mathbb{E}\,\frac{1}{X} = 1/(n-2)$.
(c) Now that we know that shrinking is a good idea, we approach the estimation from a different point of view, that of *penalisation.*
Recall Stein's setup (slide 171) : let $Y_i \sim N(\mu_i, 1)$ be independent, $i = 1, \ldots, n$. Explain why the maximum likelihood estimator $\widehat{\mu}$ can be obtained as the minimiser

$$\min_{\mu_1, \ldots, \mu_n} \sum_{i=1}^{n} (y_i - \mu_i)^2.$$

(d) We can shrink $\widehat{\mu}$ by adding a penality term that renders large values disadvantageous : for $\lambda \geq 0$ define $\tilde{\mu}_\lambda$ as the solution of

$$\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n}(y_i - \mu_i)^2 + \lambda \sum_{i=1}^{n} \mu_i^2.$$

By solving this minimisation problem, show that $\tilde{\mu}_\lambda = y/(1+\lambda)$.
(e) Find the mean squared error of $\tilde{\mu}_\lambda$ as a function of λ. *Hint :* $\mathbb{E}\, y_i - \mu_i = 0$.
(f) Show that for some values of λ, the mean squared error of $\tilde{\mu}_\lambda$ is smaller than that of $\widehat{\mu} = \tilde{\mu}_0$.
(g) Find the optimal value of λ in terms of the mean squared error. Can one use this value in practice ?
Remark. This is a particular case of ridge regression that will be seen later in the course.

Assignment 8. 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 9. ROC (Receive operating characteristic) curves are a graphical tool to perform diagnostic of a test with a binary outcome. In this exercise we will see a basic example of a ROC curve and we will discuss the choice of a suitable significance level while testing
Assume you are a laboratory offering a screening test for diabetes. A case study considered the glycemic index, measured in mg/dl, for $N = 110$ individuals between 20 and 60 years.
We assume that for healthy people the glycemic index, GLI, is distributed as a Normal with mean 80mg/dl and standard error $\sigma = 20$.
Values of GLI that exceed too much the mean might indicate that the patient suffers from diabetes.
The lab fixes the threshold $s = 120$mg/dl to signal a potential risk to the patient.
The case study finds that out of the 110 individuals studied, 10 have diabetes, but only 8 out of this 10 shows GLI measurement higher than s.
In total there are 15 patients with a GLI measurement higher than s. Out of this 15, 7 are healthy.

(i). In the following table TN = true negative, TP = true positive, FN = false negative and TP= true positive.

	healthy	desease
$GLI < s$	TN	FN
$GLI > s$	FP	TP

Discuss the previous table, in particular in relation to type I and type II error. Fill it with real numbers coming from the case study. Write down a new table where the numbers are standardized as if the two population (healthy and carrying disease) were equal.

(ii). Assume now you would like to have a screening test for a simple cold. How would you chose the critical value s ? What if you are screening for a more serious disease for which early diagnostic is crucial for recovering ?

(iii). Consider the limit cases of $s \to \infty$ and $s \to -\infty$. Write dows a table for these limit cases.

(iv). The True Positive Rate measures the proportion of positive that are correctly identified, in the notation of the table above, TPR=TP/(TP+FN) while the True Negative Rate measures the proportion of correctly identified negatives, TNR=TN/(TN+FP).
The ROC space is defined by (1-TNR) and TPR as x and y axes.
Draw on a plane the points (x, y) =(1-TNR, TPR) for the three values of s you analysed in the exercise. The curve joining the three points is an example of a ROC curve.