CS 108 - Statistics, Sections B

Fall 2019, AUA

Homework No. 09

Due time/date: Section B: 10:32 AM, 15 November, 2019

Note: Please use **R** only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1: Confidence Intervals, general

a.

Assume we have a random sample

$$X_1, ..., X_{100} \sim \mathcal{N}(\mu, 4),$$

and we want to estimate the mean μ, by constructing a CI for μ. We choose the following a Confidence Interval (CI):

$$(\overline{X} - 0.1, \overline{X} + 0.1).$$

Can we assert that this is a CI for a confidence level 95% ? Explain your reasoning.

b.

Assume we are constructing CIs for the Proportion (Parameter p in the Bernoulli Model) or the Mean of the Normal Distribution, when the Variance is known. How many times we need to increase the Sample Size to decrease the Margin of Error twice?

c.

Assume we have obtained, in some way, a 99% Confidence Level Interval Estimate

$$(1.3;\ 1.7)$$

for some Parameter θ we are estimating. What this means, how would you interpret this?

Problem 2: Constructing Confidence Intervals, Chebyshev Inequality Method

a. CI for the Exponential Model (using the second parametrization)

We consider again the Exponential Model, but with the Parameter $\theta = \frac{1}{\lambda}$: we'll write

$$X \sim \widetilde{Exp}(\theta),$$

if the PDF of X is given by

$$f(x) = \begin{cases} \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

It is easy to see that if $X \sim \widetilde{Exp}(\theta)$, then

$$\mathbb{E}(X) = \theta, \qquad Var(X) = \theta^2.$$

Now, based on a Random Sample

$$X_1, X_2, ..., X_n \sim \widetilde{Exp}(\theta),$$

and using the Chebyshev Inequality, construct a $(1 - \alpha)$-level CI for θ.

To this end, do the following steps:

- Calculate $\mathbb{E}(\overline{X})$ and $Var(\overline{X})$;
- Write the Chebyshev Inequality for $\overline{X}$ and plug the values obtained above;

Note: The right-hand side (RHS) will depend on the unknown Parameter θ - that is OK in this situation.

- Take the value of the Chebyshev Inequality ε in such a way to obtain $1 - \alpha$ on the RHS. Plug that Value of ε into the Chebyshev Inequality;
- From the obtained Inequality, obtain the CI for θ of Confidence Level $(1 - \alpha)$;
- (**R**) Take some value for θ, generate 120 times a sample of size 30 from $\widetilde{Exp}(\theta)$ (please use `rexp(30, rate = 1/theta)`), plot the corresponding CI obtained by your formula, and the actual value for θ. Use the code from our lecture slides.
- (Supplementary) Using the above CI, obtain a $(1 - \alpha)$-level CI for λ in our ordinary Exponential model $Exp(\lambda)$.

b. CI for the Uniform Model

We want to construct a CI for the Uniform Model using the Chebyshev Inequality. Assume we have a random sample

$$X_1, ..., X_n \sim Unif[0, \theta].$$

Construct a CI for θ of the confidence level $1 - \alpha$, $\alpha \in (0, 1)$, using the statistics $\hat{\theta} = 2\overline{X}$ and the Chebyshev Inequality.

(Supplementary) Write an **R** code to visualize the obtained CIs to see what proportion of CIs contain the true value of θ (like in our lecture slides).

Problem 3: Constructing One-Sided CIs, Pivoting Method

a.

Assume we have a Random Sample $X_1, ..., X_n$ from some parametric model with parameter θ. We take any function $L = L(X_1, ..., X_n, \alpha)$, and call an interval

$$(L, +\infty)$$

to be a One-Sided Upper Confidence Interval (UCI) for the parameter θ of confidence level $1-\alpha$, if

$$\mathbb{P}(L<\theta)\geq 1-\alpha.$$

Similarly, one can define the One-Sided Lower Confidence Interval (LCI) for the parameter θ of confidence level $1-\alpha$ to be an interval $(-\infty, L)$ with

$$\mathbb{P}(L>\theta)\geq 1-\alpha.$$

Construct, using the Pivoting Method, One-Sided CIs of given type for the following Parametric models:

1. $\mathcal{N}(\mu,\sigma^2)$, UCI, the parameter is μ, σ^2 is known;
2. $\mathcal{N}(\mu,\sigma^2)$, LCI, the parameter is μ, σ^2 is unknown;
3. $\mathcal{N}(\mu,\sigma^2)$, UCI, the parameter is σ^2, μ is known;
4. $\mathcal{N}(\mu,\sigma^2)$, LCI, the parameter is σ^2, μ is unknown.

(Supplementary) Write an **R** code to visualize One-Sided CIs.

Problem 4: Constructing CI in Practice and Calculating the Sample Size

a. Constructing CI

Assume that the results of some instructor's Stat course show that there are 10 A grades among 74 students. Construct a 90% confidence interval for the probability of obtaining an A grade for the Stat course of that instructor.

b. Calculating the Sample Size

Assume we want to perform a study to see which percentage of our citizens want to have "Armenian Language", "Armenian Literature" and "Armenian History" fixed by a law to be mandatory in every University curriculum. So we will choose randomly n persons and ask if they are supporting or not (they need to answer "Yes, I am for having this fixed in the law" or "No, I am not for that. These subjects need to be elective. But why the hell not *Math Reasoning* and *Descriptive Stat* courses?"). How many persons we need to ask to obtain a CI for that percentage with a Margin of Error ≤ 0.02 ?

c. Calculating the Sample Size

Assume we want to Estimate the average weight (in Kg) of all students in Armenia[1]. We assume the weights are Normally Distributed, and we assume that the Standard Deviation of weights is not larger than 5Kg. We will randomly choose n students from different universities, ask their weights, and construct a 95% CI for the average weight. How many students we need to ask to have a CI with Margin of Error $\leq$ 0.5Kg?

[1] At the beginning of the semester, say, not before the Exams ◡̈

d. (R) CI for the Normal Distribution

I have generated the following data from the Normal Distribution:

```
 [1]  1.14453951  0.33225379  4.00416672 -5.04850837  5.35766860
 [6]  0.06489233 -5.94509889 -4.03921700 -1.04401996 -1.24168001
[11] -8.06484372  1.03773668 -2.87971666 -0.71685157  0.45405920
[16]  3.30095388  0.74257551  2.13158422 -3.58842411 -2.51157687
[21] -0.05550723 -1.12422861 -4.33075098 -5.02908441 -1.91539586
[26]  1.00226791 -0.23604246 -4.60662137 -3.35289235 -3.41713159
[31] -6.73744316 -2.45756382 -1.15391642  1.50038591  3.66853943
[36] -1.05244316  4.30821758 -0.99412514  3.02116765  3.14134781
[41] -1.06686798 -5.78622173 -1.38023764 -1.87756992  5.05295459
[46] -0.62753128  0.31867113  3.79992962 -0.07876558 -5.08016167
```

- Construct, a 92% CI for μ;
- Construct, a 94.5% CI for σ^2

e. (R) Function(s) to construct CI

Write a code to construct CIs of given Confidence Level in **R**.

You can write a function for each case we have considered, say, `ci.bernoulli(x,a)`: the inputs are x - our DataSet, and a, the confidence level, and the output will be a CI for p in the $Bernoulli(p)$ Model, for the given CL. The next function will be `ci.normal.mu.givensig(x,sigma, a)`, which will calculate a CI of level a for Normal Model μ in the case when *sigma* is known, based on the Data x. Etc.

Or you can write a general function with different parameters and default parameter values: say, you can write a fiction `ci(data = x,conf.level = 0.95, ...)` with parameters

- `method = "chebyshev", "pivot", "asymptotic"` (we have not covered "asymptotic" yet, but you can later add that, and at this moment your code can print "What you want from me, I do not know it yet!" in the case if the user will use that parameter value ☺)
- `datadistrib = "bernoulli", "normalz", "normalt"`
- `parameter = "p", "mu","sigma"`
- etc.

f. (Supplementary)

1. Visit the webpage https://www.aravot.am/2018/08/18/975968/ for some (relatively old) study. Find the 95% CI for proportion of total population of persons in Armenia that think that our political parties are not doing a good job.
2. At the webpage http://armconsumer.am/news.html you can download the report about the tobacco use study in Armenia (the report is at http://armconsumer.am/pdf/report_smoking_%20free_armenia.pdf). At the pages 4-6 you can read about the sampling methodology (this is for an example, for your knowledge). In this report you will find a lot of information about the smokers in Armenia. Now, the question: based

on the very last table, page 53, Q12, estimate the percentage of people in Armenia who has a job, and construct the 90% confidence interval for that percentage.

3. We want to estimate the average price of a 3-room flat at the Davitashen district. To that end, we want to construct a 95% confidence interval for that average price. Go to https://www.list.am/en/, navigate to Real Estate -> For Sale -> Apartments, choose on the left panel the number of rooms and the location, get a sample of 10 prices, and construct the required CI. You can assume that the prices are normally distributed.

Problem 5: (Supplementary) (R)

a. $t(n)$ and $\chi^2(n)$ Distributions

- Plot the PDFs of different $t(n)$ Distributions on the same graph, taking values $n < 30$ and $n \geq 30$. Add the PDF of the Standard Normal DIstribution
- Draw a QQ-Plot for $t(3)$ vs $\mathcal{N}(0,1)$, for $t(15)$ vs $\mathcal{N}(0,1)$, for $t(30)$ vs $\mathcal{N}(0,1)$ and for $t(50)$ vs $\mathcal{N}(0,1)$. What can you infere?
- Plot the PDFs of $\chi^2(n)$ Distribution on the same graph, taking different values for n
- Simulate $\chi^2(4)$ by using the Definition (as a sum of squares of independent Standard Normals), do the Density Histogram, and add the theoretical PDF of $\chi^2(4)$. If the Sample Size is large, you need to have a good fit.