Stats exam 2014

Guillaume Filion

October 19, 2014

Please submit your answers by email to `guillaume.filion@gmail.com` before the deadline. Each question is worth 1 point. Every fraction of 24 hours passed the deadline will be penalized by 2 points. You are encouraged to work in group on these exercises. You can submit a joint answer sheet and will receive the same grade as your team members. In this case, indicate the name of all the students participating to the work.

The answers will be posted online at the address below.
`www.genomearchitecture.com/static/misc/statsexam_answers_2014.pdf`

1 Course questions

Exercise 1
What are the main steps of a test?

Exercise 2
What is the p-value of a test?

Exercise 3
What is the power of a test?

Exercise 4
How can you increase the power of a test?

Exercise 5
What is the statistic of the t-test?

2 Problem

An insecure implementation of a password-protected acess is to compare letter by letter the text enterred by the user to the real password, and to deny access as soon as a different letter is found. This is insecure because it opens a possibility for *a timing attack*, which is an attempt to hack a password by measuring the time it takes for the computer to respond.

`http://en.wikipedia.org/wiki/Timing_attack`

Suppose that my real password is `kotiki125`. If the hacker enters `abcdefg`, the insecure method will deny access immediately after comparing the first letter because `a` is different from `k`. But if the hacker enters `kotiki123`, this method will deny access upon comparing the 9-th letter, which will take roughly 9 times as long. With this information, the hacker could deduce that the first 8 letters of `kotiki123` are correct.

The idea of a timing attack is to try all the letters one by one and measure the time it takes for the computer to respond. Every time a letter matches the password, the response time will be slightly slower. By repeating the process, the password can be decrypted completely.

Exercise 6

Assume that there are 64 valid password letters (small letters, capital letters, numbers, `_` and `~`, and that you want to decrypt a password generated at random. You enter a text. What is the probability that the first letter matches the password? What is the probability that the first 3 letters match the password?

Based on previous measurements, you know that the time to compare two letters is 10 ns. In addition, for every letter comparison, there is a random overhead due to other processes running on the server. This overhead has an exponential distribution with rate 0.125 ns^{-1}.

Exercise 7

The function `rexp(8, rate=.125)` in `R` generates a random sample of size 10 drawn from an exponential distribution with rate 0.125 ns^{-1}. How to generate a random sample of size 8 representing the response time of the server when the first letter of the text does not match the password?

You have entered the text `aaaaaaaa` 8 times and have observed the following response times (in ns) 23.0, 15.7, 21.4, 12.8, 13.3, 21.3, 38.2, 15.4.

Exercise 8

Do you think that the first letter of the password is `a`? Quantify your certainty. Before you set up a complicated statistical test, look at the numbers carefully.

You know that the password starts with `wy6qU`. You now enter `wy6qUaaaaa` 8 times and obtain the following response times (in ns) 122.0, 83.8, 71.4, 136.6, 93.6, 88.1, 84.1, 123.0

In `R` you can compute the sum of 6 exponential random variables with rate 0.125 ns^{-1} by using `sum(rexp(6, rate=.125)`. Similarly, you can generate a sample of size 8 representing the response time of the server if the 6-th letter does not match by using `replicate(8,60+sum(rexp(6,rate=1/8)))`.

Exercise 9

Using this information, design a statistical test to know whether the 6-th letter of the password is `a`. Quantify your certainty.

Exercise 10

What is the power of your test?