EE 325: Probability and Random Processes

Homework on Statistics and Hypothesis Testing

1. The material for this is from parts of Sections 8.1-8.4 of the Papoulis textbook. The book has a lot more material but I have covered 8.1, most of 8.2, the MLE in Section 8.3, and some of the material before testing of distributions in Section 8.4.

2. There are several solved examples in the text. As usual, you are expected to go through them and understand them.

3. Solve problems 8.3–8.7, 8.11, 8.12, 8.21, and 8.22.

4. Estimation is also well covered in the Hajek text; see Sections 2.8, 3.7, and 4.9 for a basic understanding

EE 325: Probability and Random Processes

Programming Assignment 4: Part a

To be completed before 30 October 2024

In this assignment we will determine the premiums for a life insurance policy. The insurance scheme works as follows. Customers can buy a one-year insurance policy that covers them for the insured amount, say Ω. Each insured person will pay a premium, say α, that is determined by the amount insured and the age of the person. If there is a death of the insured the person in the year, then the family is paid Ω otherwise there is no payout to the insurer from the company. Note that the number of people that will buy the policy in a year and the number of people for whom the insured amount has to be paid are random variables.

1. Let X denote the number of insurers in a year. Assume that this is a random variable. Let Y be the number of claims in the year. We will assume that X is a Poisson random variable with mean λ and each of policy will result in a claim with probability p. Determine the joint pmf of X and Y. The surplus, denoted by S, at the end of the year is $(\alpha\mathsf{X} - \Omega\mathsf{Y})$. Clearly, the surplus is a random variable that is a function of the joint pmf of X and Y and the premium is set so that the surplus being less than some a is a low probability event. Typically, the company begins the year with a sufficient endowment and some loss, i.e., a negative surplus, may also be allowed. For a given Ω, assume you know λ, and p, the objective is to fix the premium α such that the probability of 'breaking the pot' is sufficiently low.

   Before performing calculations, determine the marginal pmf of Y.

   Assume $\Omega = 10$. We will consider two values of λ—10,000 and 100,000, and two values of p—0.01 and 0.02. For $a = 10, 0, -10, -20, -30, -100$, find the minimum α such that $\mathsf{Pr}(\mathsf{S} < a) < 0.01$. Tabulate your results for the four combinations of (λ, p) and the six values of a.

   Rather than using ratios of factorials of large numbers, you can use Gaussian approximation suitably in the probability formula that you calculated above. Specifically, note that Gaussian approximations for the Poisson and the binomial distributions are available; the latter was done in class. These can perhaps be used suitably in the formulas that you have derived to compute the joint pmf.

2. Let us take the same situation as above. except that there are two types of policies targeted at two different demographics with (Ω_i, α_i) targeting a population demographic characteristic of $(\lambda_i, p_i,)$ for $i = 1, 2$. Here the insurance company is willing to do some cross subsidy. Here we can consider three surpluses—the surpluses from each type of policy, denoted by S_i and the total surplus, S and want that $\mathsf{Pr}(\mathsf{S}_i < a_i) < 0.01$

for $I = 1, 2$, and $\mathsf{Pr}(\mathsf{S} < a) < 0.01$. Determine α_1 and α_2 for the following parameters: $\lambda_1 = 100,000$, $\lambda_2 = 5,000$, $\Omega_1 = 10$, $\Omega_2 = 100$ $p_1 = 0.02$, $p_2 = 0.02$, $a_1 = -100$, $a_2 = -15$ and $a = 105$.

3. You were to be ready with the programs for the above two parts before 29 October. In **Part B** of this assignment, we will be working (synthetic) data. You are given the data for the number of policies sold and the number of claims made in the preceding N years. Using this data, and making the strong assumption that that the conditions have not changed in these N years, write a program to obtain the point estimate and the 95% confidence interval for λ and p.

   (a) The app available at the URL `https://karrthik-arya-ee325-ass4-generator-otsy8d.streamlit.app/` can be used to generate N samples of policies sold and claims made. Use the output from this app in the estimation programs you have written. Compare your estimates with the true values that you have given as input to the app. Repeating this a few times will give you a sense of how estimates work.

   (b) Using $\lambda = 100,000$ and $p = 0.02$ generate $N = 20$ samples from the app and obtain the 95% confidence intervals for λ and p. Using these interval estimates redo the table from the first part to obtain the α. Repeat for $N = 100$ samples.

Discussion

1. In both the parts you are essentially searching for the 'right α.' And you will need to perform summations over a large number of indexes. Using the numerical insights discussed in class, and developing more of those insights on your own, you can suitably reduce the search space.

2. For the second part above several values of α_1, α_2 would be possible. Which of those would you choose?

3. What do you think "cross subsidy" means? How would you achieve it? Argue for and against cross subsidy in this case.

4. If, instead of using the joint pmf of X and Y above, you had assumed that they were independent would your answers for the two parts change substantially. Explain.

Sumbission

1. Derivation of the marginal pmf of Y.

2. The code and the table for the first part.

3. The code and the values of α_1 and α_2 for the second part.
4. Code for the point and interval estimates in Part B.
5. The sample data that you obtained and the two tables for Part B.
6. A short writeup jointly addressing items 2 and 3 in the discussion.