Maths and Statistics Foundations for Analytics, Autumn 2020 Imperial College Business School
Zhe Liu

Assignment 3: Statistics

Total Points: 100 Due: 11:59pm 20 Oct, 2020

Rules:

1. This is a group assignment. Within each group **I strongly encourage everyone to attempt each question by his/herself first** before discussing it with other members of the group.
2. You are recommended to **type out your solution** and submit a PDF file. If you choose to write it down and submit a scanned version, please ensure clear handwriting.
3. `R` is the default package / programming language for this course so you should use `R` for any programming questions in this assignment.

---

1. **Geometric Distribution and MLE (40 points)**

   Consider a sample $x_1, x_2, \ldots, x_N$ of i.i.d. random variables where each $X_i = x_i$ has a geometric distribution with unknown parameter θ so that for each i

   $$\mathrm{P}(X_i = x_i) = \theta(1-\theta)^{x_i - 1}, \quad x_i = 1, 2, 3, \ldots$$

   (Recall that the geometric distribution can be interpreted in two ways, here we think of it as the distribution of how many coin tosses are *needed* until the first "head", where the coin tosses are i.i.d. with P("head") $= \theta$.) Given the sample $X_1, X_2, \ldots, X_N$ the goal is to obtain the maximum likelihood estimator (MLE) of θ.

   (a) Give an expression for the likelihood $L(\theta)$. (10 points)

   (b) Give an expression for the log-likelihood $l(\theta)$. (10 points)

   (c) Maximize $l(\theta)$ to obtain the MLE estimator $\widehat{\theta}_{\mathrm{MLE}}$. (10 points)

   (d) Does your estimator $\widehat{\theta}_{\mathrm{MLE}}$ make intuitive sense? Justify your answer. (10 points)

---

2. **Weldon's Dice Data (20 points)**

   An English biologist named Weldon was interested in the "pip effect" in dice – the idea that the spots or "pips", which on some dice are produced by cutting small holes in the surface, make the sides with more spots lighter and more likely to turn up. Weldon threw 12 dice 26,306 times for a total of 315,672 throws and observed that a 5 or 6 came up on 106,602 throws, i.e. 33.77% of the time.

(a) Construct an approximate 95% confidence interval for p, the true probability of obtaining a 5 or a 6. (15 points)

(b) Based on the approximate CI you obtained in part (a), what can you tell about the "pip effect" or fairness of these dice? (5 points)

---

3. **Hypothesis Testing (40 points)**

(a) Define appropriate null and alternative hypotheses in the following case: Oxbridge University announces that on average it takes only 3 months for its MBA graduates to get a new job. A higher education institute that ranks MBA programs thinks that it would take longer than Oxbridge's claim. (10 points)

(b) We would like to test the hypothesis $H_0 : \mu \geq 6$ against $H_1 : \mu < 6$. The sample mean $\bar{X} = 3$, sample variance $s^2 = 9$, and sample size $n = 100$. At 5% significance level, can we reject the null hypothesis? Please explain your answer. (10 points)

(c) We would like to test the hypothesis $H_0 : \mu \geq 3$ against $H_1 : \mu < 3$. The sample mean $\bar{X} = 1$ and sample variance $s^2 = 225$. What is the minimum sample size for us to be able to reject the null hypothesis at 5% significance level. Please explain your answer. (10 points)

(d) We would like to test the hypothesis $H_0 : \mu = 3$ against $H_1 : \mu \neq 3$. The sample mean $\bar{X} = 2.8$, sample variance $s^2 = 1$, and sample size $n = 100$. Please derive the p-value and state whether or not we can reject the null hypothesis at 5% significance level. (10 points)