

Part I: Theory and Calculation

1. The scores in a class quiz are as follows: 58, 62, 62, 63, 65, 65, 65, 68, 69, 72, 72, 75, 76, 78, 79, 81, 84, 84, 85, 92, 94, 95, 98.
 - a) What is the mean score in the quiz?
 - b) What is the median score in the quiz?
 - c) What is the mode of the quiz?
2. A random sample of size 81 is taken from a population that has a mean of 24 and variance 324. Use the central limit theorem to determine the probability that the sample mean lies between 23.9 and 24.2.
3. A box contains a mix of red and blue balls whose exact composition of red and blue balls is not known. If we draw 20 balls from the box with replacement and obtain 12 red balls, what is the maximum-likelihood estimate of p , the probability of drawing a red ball?
4. A company claims that the boxes of detergent that it sells contain more than the current 500 grams of detergent each. From past experience the company knows that the amount of detergent in the boxes is normally distributed with a standard deviation of 75 grams. A worker takes a random sample of 100 boxes and finds that the average amount of detergent in a box is 510 grams. Test the company's claim at the 0.05 level of significance.
5. Nefeli, a student in a probability class, takes a multiple-choice test with 10 questions and 3 choices per question. For each question, there are two equally likely possibilities, independent of other questions: either she knows the answer, in which case she answers the question correctly, or else she guesses the answer with probability of success $1/3$.
 - a) Given that Nefeli answered correctly the first question, what is the probability that she knew the answer to that question?
 - b) Given that Nefeli answered correctly 6 out of the 10 questions, what is the posterior PMF of the number of questions of which she knew the answer?
6. Students in a probability class take a multiple-choice test with 10 questions and 3 choices per question. A student who knows the answer to a question will answer it correctly, while a student that does not will guess the answer with probability of success $1/3$. Each student is equally likely to belong to one of three categories, $i = 1, 2, 3$: those who know the answer to each questions with corresponding probabilities b_i , where $b_1 = 0.3$, $b_2 = 0.7$ and $b_3 = 0.95$ (independent of other questions). Suppose that a randomly chosen student answers k questions correctly.
 - a) For each possible value of k , derive the MAP estimate of the category that this student belongs to.

- b) Let M be the number of questions that the student knows how to answer. Derive the posterior PMF, and the MAP and LMS estimates of M given that the student answered correctly 5 questions.
7. A police radar always overestimates the speed of incoming cars by an amount that is uniformly distributed between 0 and 5 miles/hour. Assume that car speeds are uniformly distributed between 55 and 75 miles/hour. What is the LMS estimate of car's speed based on the radar's measurement?
8. Let Θ be a positive random variable, with known mean μ and variance σ^2 , to be estimated on the basis of a measurement X of the form $X = \sqrt{\Theta}W$. We assume that W is independent of Θ with zero mean, unit variance, and known fourth moment $E[W^4]$. Thus, the conditional mean and variance of X given Θ are 0 and Θ , respectively, so we are essentially trying to estimate the variance of X given an observed value. Find the linear LMS estimator of Θ based on X , and the linear LMS estimator of Θ based on X^2 .

Part II: Simulation

Guidelines and rules for this part (please read carefully):

- For each problem, you are expected to first do necessary mathematical analysis (key words: 'calculate') of the problem, then use simulation to verify your analysis results.
 - For mathematical part, please show process of the solution in the report. Answers with only results will not earn full grade.
 - For the simulation part, in your report, please clearly state how you set up the simulations, your ideas and the necessary intermediate results and figures/tables you got from simulations. You will use MATLAB to run the computer simulations. The codes should run without problems on the TA's computer.
 - For each problem, please turn in: (i) source codes, and (ii) a written report to the Gradescope.
 - Penalty for late submission: 10% penalty if late within one day; Not acceptable (0 grade) if late more than one day.
 - You may discuss coding/debugging issues with your classmates. You cannot, however, share code. It is a violation of ethical policies if you request code from someone else, or if you give code to someone else.
1. Given two geometric RVs $X \sim \text{geom}(p)$, $Y \sim \text{geom}(p)$, and X and Y are independent, show that the PMF of $Z = X + Y$ is given by

$$p_Z[k] = p^2(k-1)(1-p)^{k-2}, \quad k = 2, 3, \dots$$

To avoid errors, use the discrete unit step sequence. Next, for $p = 1/2$ generate realizations of Z by first generating realizations of X , then generating realizations of Y and adding each pair of realizations together. Estimate the PMF of Z and compare it to the true PMF.

2. If $X_i \sim \mathcal{N}(1,1), i = 1, 2, \dots, N$ are independent identically distributed (I.I.D.) random variables, plot a realizations of the sample mean random variable versus N . Should the realization converge, and if so, to what value?