ASDS - Applied Statistics with R

Fall 2018, YSU

Homework No. 06

Due time/date: 21:20, 20 November, 2018

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. Assume $X_1, X_2, ...$ is a sequence of IID random variables with the distribution

X_k	-1	1
$\mathbb{P}(X_k = x)$	0.5	0.5

Assume also $X_0 \equiv 0$, and denote

$$Y_n = X_0 + X_1 + X_2 + ... + X_n, \qquad n \in \mathbb{N} \cup \{0\}.$$

The sequence $Y_0, Y_1, Y_2, ...$ is called a **1D random walk**: imagine a drunk man standing at the point 0 at time $t = 0$ (the initial position, say, home). At the next time instant $t = 1$, he goes randomly either 1 units to the left or 1 units to the right with probabilities 0.5, and Y_1 is the position of our drunk man at time $t = 1$ (X_1 is $+1$ if he chooses to go right, and is -1, if he chooses to go to left, and $Y_1 = Y_0 + X_1$ is his new position). At time $t = 2$, he goes randomly to the left or right 1 units randomly, with equal probabilities, and his position on the real line at time $t = 2$ is Y_2, and so on. That is, Y_n is a r.v. showing possible positions of our drunk man at time n.

a. What is the set of all possible values of Y_n?

b. Give the PMF of Y_2;

c. Calculate the expected position of our drunk man at time $t = n$, and the variance of Y_n;

d. Approximate, for a large n, the probability that our drunk man will be between the points a and b, i.e., approximate $\mathbb{P}(a \leq Y_n \leq b)$;

e. (**R**) Draw some possible paths for that man, with different colors.

f. (Supplementary) Calculate the probability that $Y_n = 0$, i.e., at the time $t = n$, our drunk man will return to the initial position (home).

g. (Supplementary) Prove that along the time, our drunk man will return to the initial position (home) infinitely many times.

Problem 2. Assume $X_k \sim Unif[0, 1]$, $k \in \mathbb{N}$, are IID.

a. Calculate[1] $\lim\limits_{n\to+\infty} \dfrac{X_1+X_2+...+X_n}{n}$;

b. Assume $g \in C(\mathbb{R})$. Calculate the limit (in terms of g)

$$\lim_{n\to+\infty} \frac{g(X_1)+g(X_2)+...+g(X_n)}{n}$$

c. Calculate the limit

$$\lim_{n\to+\infty} \frac{X_1^2+X_2^2+...+X_n^2}{n}$$

d. Calculate the limit

$$\lim_{n\to+\infty} \sqrt[n]{(1+X_1)\cdot(1+X_2)\cdot\ldots\cdot(1+X_n)}$$

e. (Supplementary) Using *R* or MatLab or any other Math software, calculate, using the Monte Carlo Method, the integrals

$$\int_0^1 \sin(x)dx, \qquad \int_0^1 \frac{dx}{1+x^2}, \qquad \iint_{x^2+y^2\le 1} \frac{xy+\sin(x^2+y^3)}{e^x+e^y+1}dxdy,$$

and compare with the actual values for the first two integrals.

Problem 3. Assume X_k are IID r.v. with $X_k \sim \mathcal{N}(\mu,\sigma^2)$. Calculate the limit

$$\lim_{n\to+\infty} \frac{X_1+X_2+...+X_n}{X_1^2+X_2^2+...+X_n^2}.$$

Problem 4. Assume X_k are IID r.v. with mean μ and variance σ^2. Calculate the limit

$$\lim_{n\to+\infty} \frac{(X_1-\mu)^2+(X_2-\mu)^2+...+(X_n-\mu)^2}{n}.$$

Problem 5. Assume we are tossing a fair coin 10000 times. What is the (approximate) probability that the number of heads shown will be in between 4950 and 5150? You are free to calculate the Standard Normal r.v. probabilities with a computer software (i.e., if $Z \sim \mathcal{N}(0,1)$, you can calculate $\mathbb{P}(a < Z < b)$ using, say, MatLab or R).

Problem 6. Assume we are throwing a die 80 times. Let X be the number of 5's shown, and Y be the total sum of numbers shown.

a. Calculate $\mathbb{E}(X)$, $Var(X)$, $\mathbb{E}(Y)$, $Var(Y)$;

b. Calculate approximately

$$\mathbb{P}(20 \le X \le 25) \qquad \text{and} \qquad \mathbb{P}(200 \le Y \le 220).$$

c. (Supplementary) Read about the Continuity Correction, and do the part **b.** using the Continuity Correction.

---

[1] Here everywhere we use limits in the almost surely convergence sense, i.e., by writing $\lim Y_n = Y$ we mean that $Y_n \to Y$ almost surely.

Problem 7. [2] In any given day, a certain email account gets a number of spam emails that has a Poisson distribution with mean 200. What is the approximate probability that it receives less than 190 spam emails in a day?

Problem 8. How many times do you need to roll a die to be at least $\approx 99\%$ certain that the sample mean is between 3 and 4?

Problem 9. A multiple-choice test has 100 questions, each with four alternatives. At least 80 correct answers are required for a passing grade. On each question, you know the correct answer with probability $3/4$, otherwise you guess at random. What is the (approximate) probability that you pass?

[2]The problems no. 6-9 are from the book by P. Olofsson and M. Andersson, *Probability, Statistics and Stochastic Processes*, 2nd Edition.