بسمه تعالی

تمرین شبیه سازی اول درس یادگیری ماشین آماری **بخش ۱، احتمالات**

فرمت گزارش:

گزارش باید به زبان فارسی و در قالب فایل PDF باشد. در گزارش تحلیل و نتیجه‌گیری خود را در رابطه با هر تمرین به طور مختصر، در حد یک پاراگراف، بیان نمایید.

فایل گزارش خود را به شکل Report1_StdNum.pdf نامگذاری نمایید(مانند Report1_8931064.pdf)

فرمت کدها:

برای هر تمرین باید فایل کد جداگانه در محیط MATLAB ، R یا Python تهیه شود. هر فایل کد خود را به شکل CE1_ProblemNum_StdNum نامگذاری کنید.

نحوه تحویل:

فایل‌های کد و گزارش خود را که طبق فرمت‌های فوق تهیه شده‌اند، در قالب یک فایل فشرده در سایت درس بارگذاری نمایید.

فایل فشرده را به شکل CE1_StdNum نامگذاری نمایید.

مهلت ارسال تمرین ساعت 23:55 دقیقه‌ی روز دوشنبه مورخ ٦ آذر ماه می‌باشد.

ضمنا به ازای هر روز تاخیر در ارسال تمرین، 10 درصد از نمره‌ی آن کم می شود.

هر گونه سوال در مورد تمرین را میتوانید از طریق ایمیل به آدرس z_naraghi@aut.ac.ir ارسال نمایید.

---

1. Suppose a coin has probability p of falling heads. If we flip coin many times, we would expect the proportion of heads to be near p. Take p=0.4 and n=1000 and simulate n coin flips. Plot the proportion of heads as a function of n. Repeat for p=0.04.

2. Suppose we flip a coin n times and let p denote the probability of heads. Let X be the number of heads. Intuition suggests that X will be close to $n\,p$. To see if this is true, we can repeat this experiment many times and average the X values. Carry out a simulation and compare the average of the X's to $n\,p$. Try this for p=0.4 and n=10, 100, 1000.

3. Use computer simulation to show that changing selected door in "Monty Hall Problem" leads to higher winning probability. Simulate game for 2000 rounds and change the door in first 1000 rounds. Calculate winning probability in two conditions and compare them.
   - If you are not familiar with "Monty Hall problem", read problem 2 of chapter 2 in course textbook.

4. Let $X{\sim}N(5,18)$. Solve the following parts using computer simulation.
   (a) Find $P(X{<}9)$.
   (b) Find $P(X{>}{-3})$.
   (c) Find x such that $P(X{>}x){=}0.05$.
   (d) Find $P(0{\le}X{<}4)$.
   (e) Find x such that $P(|X|{>}|x|){=}0.05$.

5. Determine the probability that in a class of 28 students two or more students have birthdays on January 1. Use a computer simulation to verify your result.

6. Consider tossing a fair die. Let $A{=}\{2,3,6\}$ and $B{=}\{1,2,3,4\}$. Simulate draws from the sample space and verify that $P(AB){=}P(A)P(B)$. Now consider tow **disjoint** events A and B that are not independent. Compute $P(A)$, $P(B)$ and $P(AB)$. Compare the calculated values to their theoretical values. Draw van diagram and report the results.

7. (a) Generate random samples for rolling a weighted dice, where the probability to get the value 6 is 1/4 and the rest of the values are equally probable.
   - First, roll the dice for $N{=}10$ times (hint: randi or rand function)
   - Plot the normalized histogram (i.e., the pdf estimate)

   (b) Define the true pdf for the weighted dice and plot it on top of the estimated pdf besides using the plot-function, try also
   - stairs (hint: stairs)
   - stem (hint: stem)

8. Let $X{\sim}N(0,1)$ and let $\mathrm{Y}{=}e^X$:
   (a) Find the pdf for Y. Plot it.
   (b) Generate a vector $x{=}(x_1,x_2,\ldots,x_{10000})$ consisting of 10000 random standard Normals. Let $y{=}(y_1,y_2,\ldots.,y_{10000})$ where $y_i{=}e^{x_i}$. Draw a histogram of y and compare it to the PDF you found in part (a).

9. Let $X_1,X_2,\ldots,X_n$ be $N(0,1)$ random variables and $\bar{\mathrm{X}}_n$ is sample mean of first n samples. Plot $\bar{X}_n$ versus n for $=1,\ldots,10000$. Repeat for $X_1,X_2,\ldots,X_n{\sim}Cauchy$. Explain why there is such a difference.

10. Let $X_1,\ldots,X_n$ be IID with mean μ and variance $\sigma 2$. $\bar{X}_n$ is the mean of n samples. Then $\bar{X}_n$ is a statistic, that is a function of the data. Since $\bar{X}_n$ is a random variable, it has a distribution. This distribution is called the *sampling distribution of the statistic.* Recall that $E(\bar{X}_n){=}\mu$ and $(\bar{X}_n){=}\ \sigma^2/n$. Don't confuse the distribution of the data f_x and the distribution of the statistic $f_{\bar{X}_n}$. To make this clear, let $X_1,\ldots,X_n{\sim}Uniform(0,1)$. Let f_x be the density of $Uniform(0,1)$. Plot f_x. Find $E(\bar{X}_n)$ and $(\bar{X}_n)$. Plot them as a function of n. Now simulate the distribution of $\bar{X}_n$ for $n{=}1,\ 10,\ 30,\ 100$. Check that

the simulated values of $E(\bar{X}_n)$ and $V(\bar{X}_n)$ agree with your theoretical calculations. What do you notice about the sampling distribution of $\bar{X}_n$ as n increases?

11. Use a computer simulation to generate realizations of a $Poisson$ (6) random variable by approximating it with a $binomial$ (100, 0.06) random variable. What is the average value of X?

12. A large circular dartboard is set up with a "bullseye" at the center of the circle, which is at the coordinate (0,0). A dart is thrown at the center but lands at (X,Y), where X and Y are two different Gaussian random variables. What is the average distance of the dart from the bullseye? (Use computer simulation.)

13. Solve the following parts using computer simulation and verify your result theoretically.
(a) Estimate the PDF of $X=Y_1-Y_2$, where Y_1 and Y_2 are uniform random variables. What is the most probable range of values?
(b) Estimate the PDF of $X=Y_1Y_2$, where Y_1 and Y_2 are uniform random variables. What is the most probable range of values?

14. Write a program to generate a pair of Gaussian random numbers $(X1,X2)$ with zero mean and covariance $(X_1^2) = 1$, $E(X_2^2)=\frac{1}{3}$, $E(X_1X_2)=\frac{1}{2}$. Generate 1000 pairs of such numbers, evaluate their sample averages and sample covariance

15. (a) Generate 1000 samples to estimate $P(|\bar{X}-p|>\epsilon)$ in Example 5.3 of textbook for $n=100$, $\epsilon=0.2$ and $p=0.3$. Estimate $\bar{X}$ after each $n=100$ sample generation. Compare the result with boundaries in Example 5.3, Example 5.6 of course textbook.
(b) Repeat part (a) for $p=0.5$ and report results.

16. (Computer Simulation) Consider X1~ Binomial(1000, 0.3), X2~ Binomial(1000,0.5) and X3~ Binomial(2000,0.5). Show X1+X2 and X2+X3 have binomial distributions and find their parameters.

17. (Computer Simulation) Consider X_i~N(0, $\frac{1}{i}$). Draw CDF of X_i for i=1, 100, 1000, 10000. Consider an arbitrary very small number (epsilon). Show X_i converges to 0 in probability and distribution.