HW 1 - due 01/22 at 11:59 pm.

Math 185, Fall 25, Rava

Follow closely the 'Hw guide' under Files in the folder 'Course Contents' on how to write, scan and submit your homework.
On any problem involving R, you should include your code and output as part of your answer. You may take a screenshot of the code/output, use R Markdown, or simply write it by hand.
Be careful with notation, remember to define the parameters and the random variables you intend to use.

1 Exercise 1

Consider a discrete random variable X with the following distribution:

$$P(X=1)=0.3, \quad P(X=2)=0.5, \quad P(X=3)=0.15, \quad P(X=4)=0.05$$

a) [2 points] Find the expected value $E(X)$.

b) [2 points] Find $P(X \leq 2)$.

c) [6 points] Implement in R a function 'simulation(n)' that, given as argument the sample size n, return a sequence of n random elements from this distribution.

d) [2 points] Use 1000 simulations from this distribution to estimate the mean of the distribution. Compare it with the expected value found in part a).

e) [2 points] Use 1000 simulations from this distribution to estimate $P(X \leq 2)$. Compare it with $P(X \leq 2)$ found in part b).

2 Exercise 2

Let $X \sim B(n,p)$ and $\hat{p} = X/n$. We learn in class that $\hat{p} \sim_{approx} N\left(p, \frac{p(1-p)}{n}\right)$ when n is large enough. We learned that the approximation works well when $0 < p - 3\sqrt{p(1-p)/n} < p + 3\sqrt{p(1-p)/n} < 1$. In this exercise, we use simulations to check this claim.

a) [4 points] Consider $n = 100$ and $p = 0.6$. Simulate in R 1000 realizations of $\hat{p}$. Plot the histogram of the 1000 realizations. Does the distribution look normal? If yes, does it make sense? If not, does it make sense? (You can use the built-in function 'rbinom' to perform simulations).

b) [4 points] Consider $n = 100$ and $p = 0.015$. Simulate in R 1000 realizations of $\hat{p}$. Plot the histogram of the 1000 realizations. Does the distribution look normal? If yes, does it make sense? If not, does it make sense? (You can use the built-in function 'rbinom' to perform simulations).

3 Exercise 3

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be an i.i.d sample. Suppose you are interested in making inference for the variance σ^2. You can uses the following formula to construct a $(1-\alpha)\%$ CI for σ^2: $\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right]$, where $\chi^2_{\alpha/2,n-1}$ is the $\alpha/2$ quantile of the chi-squared distribution with $n-1$ degrees of freedom (in R: $qchisq(\alpha/2, n-1)$) and S^2 is the sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ (in R: $var(x)$). To test $H_0 : \sigma^2 = \sigma_0^2$, you can use the following test statistic $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$, that has a chi-squared distribution with $n-1$ degrees of freedom under the null. For a two-sided alternative, you reject the null if χ^2 is either $\leq \chi^2_{\alpha/2,n-1}$ or $\geq \chi^2_{1-\alpha/2,n-1}$. In this exercise we use simulations to check the actual coverage of the interval and the actual type I error rate of the test under different scenarios. In the following you can use the R built-in functions to simulate RVs.

a) [4 points] Let $n = 100$, $\mu = 1$, and $\sigma^2 = 1$. Use 1000 simulations to estimate the coverage of the procedure. Is this close to the nominal coverage? Briefly comment.

b) [4 points] Repeat exercise a) under the following three different scenarios:

- $X_1, \ldots, X_{10} \sim N(1, 1)$
- $X_1, \ldots, X_{100} \sim Exp(1)$
- $X_1, \ldots, X_{10} \sim Exp(1)$

Comment on your results. (Try to address the following questions: Does it seem that the nominal coverage is obtained? If yes explain why you think it is the case, if not explain why you think it is the case. Does a big sample size make a difference in obtaining the nominal coverage? If yes, does it make sense? If not, does it make sense?)

c) [4 points] Let $n = 100$, $\mu = 1$, and $\sigma^2 = 1$. Use 1000 simulations to estimate the probability of type I error rate of the test. Is this close to the probability of type I error that the test should have? Briefly comment.

d) [4 points] Repeat exercise c) under the following three different scenarios:

- $X_1, \ldots, X_{10} \sim N(1, 1)$
- $X_1, \ldots, X_{100} \sim Exp(1)$
- $X_1, \ldots, X_{10} \sim Exp(1)$

Comment on your results. (Try to address the following questions: Does it seem that the estimated probability of type I error rate is close to the probability of type I error that the test should have? If yes explain why you think it is the case, if not explain why you think it is the case. Does a big sample size make a difference? If yes, does it make sense? If not, does it make sense?)