COMP 1433 Quiz 2 (Tuesday)

1. Which of the following R function allows us to draw a random number from a normal distribution?
   A. rnorm(1)
   B. dnorm(1)
   C. pnorm(1)
   D. qnorm(1)

   rnorm(n, mean, sd)

   n: Number of obervations

   mean: Mean value for the sample data. Its default value is zero.

   Std: Standard deviation. Its default value is 1.

   data <- rnorm(10)

2. (True or False) Fixing the random seeds in R programing before calling a random number generator is crucial to ensure reproducibility of a simulation.
   A. True
   B. False

   Exclude the randomness, make sure everything is deterministic for reproducibility.

3. (True or False) Monte Carlo simulation is based on the law of large numbers.
   A. True
   B. False

   Monte Carlo Principal: In repeated independent tests with the same actual probability *p* of a particular outcome in each test, the chance that the fraction of times that outcome occurs differs from p converges to zero as the number of trials goes to infinity.

   Law of large numbers: the average of the results obtained from a large number of trials should be close to the expected value and tends to become closer to the expected value as more trials are performed.

4. (True or False) In a queueing system, random variable t indicates the time between two successive customers' arrival. Then, t usually satisfies Possion distribution.
   A. True
   B. False

   Let the time between successive arrivals into some system be exponentially distributed, and let N be the number of arrivals in a fixed interval of time of length t. Then N (a discrete random variable) has the Poisson distribution, and

$$\Pr(N = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$

5. Which of the following about Monte Carlo simulation is NOT true?
   A. It is a very effective method possibly allow 100% accuracy.
   B. The confidence of the estimation largely depends on the variance of samples.
   C. Larger sample size may be helpful to draw unbiased results.
   D. The training method of Naïve Bayes can be seen as the Monte Carlo simulation.

P17 Lecture 9

B:

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

converges to the expected value:

$$\overline{X}_n \to \mu \quad \text{as } n \to \infty.$$

Therefore, with larger sample size, the estimation will be more reliable.

C:

*(***Chebyshev's inequality***)*

$$\mathbf{P}(|X| \geq a) \leq \frac{\mathbf{E}[X^2]}{a^2}$$

$$\mathbf{P}(|Y_n - Y| > \epsilon) \leq \frac{\mathbf{E}[(Y_n - Y)^2]}{\epsilon^2}$$

$\mathbf{E}[(Y_n - Y)^2]$ is the variance for the observation Yn. Therefore, with smaller variance, Yn will have smaller chance that diverge from center Y

D:

Markov Chain Monte Carlois a class of techniques for sampling from a probability distribution and can be used to estimate the distribution of parameters given a set of observations.