Problem Set 6

1. What does the Law of Large Numbers state?
   (a) The sum of independent and identically distributed (iid) random variables is always normally distributed.
   (b) As the sample size increases, the sample mean will converge to the population mean.
   (c) The probability of any event approaches 1 as the number of trials increases.
   (d) The sample variance decreases as the sample size increases.
2. For the CLT to hold, which of the following must be true about the sample?
   (a) The sample must come from a normal distribution.
   (b) The sample must be large enough, regardless of the population distribution.
   (c) The sample mean must be equal to the population mean.
   (d) The sample variance must be equal to the population variance.
3. Let $\{X_1, X_2, \ldots, X_n\}$ be an i.i.d sample from the population. Which of the following is true accoring to the Central Limit Theorem?
   (a) $X_1 + \cdots + X_n$ is normally distributed for large n
   (b) X_n becomes normally distributed for large n
   (c) The sample mean $\bar{X}_n$ equals the true mean for lage n
   (d) None of the above
4. Which of the following can be a sample?
   (a) A group of 100 students selected from a school of 1,000 students
   (b) All the population in China
   (c) All the countries in the world
   (d) All of the above
5. Which of the following statements about population parameters and sample statistics is true?
   (a) Population parameters are fixed, while sample statistics vary from sample to sample.
   (b) Both the population parameters and sample statistics are calculated from data.
   (c) Sample statistics are always smaller than population parameters.
   (d) None of the above

6. Which of the following is true about the sampling distribution?
   (a) It is the distribution of the conponents in the sample
   (b) It is the distribution of an estimator
   (c) It is a subset of the population distribution
   (d) None of the above
7. Which of the following statements is true about a statistic?
   (a) It is a function of the random sample
   (b) It is the estimate of some population parameter
   (c) It is a random variable
   (d) All of the above
8. Which of the following statements is true about the sampling distribution of the sample mean $\bar{X}$?
   (a) It is always normally distributed, regardless of the sample size.
   (b) Its standard deviation decreases as the sample size increases.
   (c) It approaches the population distribution as sample size increases.
   (d) All of the above.
9. Don, Joe, and n other guests arrive at a party at i.i.d times drawn from a Uniform distribution with support $[0, 1]$, and stay until the end (time 0 is the party's start time and time 1 is the end time). Let T_D, T_J denote the arriving time of Don and Joe respectively.
   (a) Find $P(T_D < T_j < T_J)$ assuming $T_D < T_J$ and T_j is the arriving time for a random guest j.
   (b) On average, how many of the n other people will arrive between Don and Joe?
   (c) Jim and Tim are two of the other guests. Determine whether the event "Jim arrives between Don and Joe" is independent of the event "Tim arrives between Don and Joe".

Solutions

1. B
2. B
3. A
4. D
5. A
6. B
7. D
8. B
9. Solution:

   (a) Guest j arriving between Don and Joe corresponds to the event $T_j \in |T_D - T_J|$. The probability of T_j being in the interval is equivalent to the expected interval length $E|T_D - T_J|$ (relative to the total length 1). To simplify the notation, we use X, Y to represent T_D, T_J:

$$\begin{aligned} E|T_D - T_J| &= \int\int_{x<y} (y-x)dxdy + \int\int_{x>y} (x-y)dxdy \\ &= 2\int\int_{x<y} (y-x)dxdy \\ &= 2\int_0^1 \int_0^y (y-x)dxdy \\ &= \frac{1}{3}. \end{aligned}$$

   (b) By symmetry, the expected number of people out of n arriving between Don and Joe time is $n/3$.

   (c) Jim and Tim's arriving time are conditionally independent given the interval $|T_D - T_J|$. Let I denote the interval.

$$P(T_{Jim} \in I \cap T_{Tim} \in I|I) = P(T_{Jim} \in I|I)P(T_{Tim} \in I|I) = I^2;$$

$$P(T_{Jim} \in I \cap T_{Tim} \in I) = \int_I I^2 P(I) = \int_0^1 \int_0^1 (y-x)^2 dxdy = \frac{1}{6}.$$

   So $P(T_{Jim} \in I \cap T_{Tim} \in I) \neq P(T_{Jim} \in I)P(T_{Tim} \in I)$ as $P(T_{Jim} \in I) = P(T_{Tim} \in I) = \frac{1}{3}$. The two event are not independent. The intuition is that, despite they are conditionally independent given the interval, they both dependent the interval which is a random variable. Jim arrives within the interval us something about the interval (perhaps the interval is wide), which in turns affects the probability that Tim arrives within the interval.