STAT 757 – HW #2 – SOLUTIONS

1. Suppose random vector $\mathbf{X} = (X_1, X_2, X_3)$ is a random sample of three observatons from a discrete uniform distribution with sample space $S = \{1, 2, 3\}$. Describe the set of outcomes (i.e., write out the full set of one or more triplets) that corresponds to the event

$$\sum_{i=1}^{3} \mathbf{X_i} = 5.$$

Answer: All triplets that sum to 5, i.e., $\{113, 131, 311, 122, 212, 221\}$.

2. Suppose events A and B have $P(A) = 0.2$ and $P(B) = 0.25$. What is the probability of A or B occurring if the two events are mutually exclusive? If $P(A \cup B) = 0.4$ are A and B mutually exclusive and/or independent? If C is the complement of $A \cup B$, what is $P(C)$?

Answer: Since $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - 0 = 0.45$. If $P(A \cup B) = 0.4$, then $P(A \cap B) = 0.05$, and the two sets are not mutually exclusive. Since $P(A)P(B) = 0.05$, A and B are, by definition, independent. Lastly, $P(C) = P((A \cup B)')) = 1 - P(A \cup B) = 0.6$.

3. Suppose you draw 3 cards from a standard deck of 52, and all three are Aces. What is the probability of the fourth card also being an Ace, i.e., what is P(4th an Ace | 3 Aces)? Answer this question (1) from a counting-based argument, and (2) using the definition of conditional probability.

Answer: Using the definition of conditional probability,

$$P(\text{4th an Ace} \mid \text{3 Aces}) = P(\text{4 Aces} \cap \text{3 Aces})/P(\text{3 Aces}) = P(\text{4 Aces})/P(\text{3 Aces}).$$

Since $P(\text{4 Aces}) = 1/\binom{52}{4}$, and $P(\text{3 Aces}) = 4/\binom{52}{3}$ (there are 4 distinct sets of 3 aces), we have $P(\text{4th an Ace} \mid \text{3 Aces}) = P(\text{4 Aces})/P(\text{3 Aces}) = \binom{52}{3}/\binom{52}{4}/4 = \mathbf{1/49}$

One could also argue that this is equivalent to the probability of drawing the 1 remaining Ace from the remaining 49 cards, and arrive at the same answer. (However, it's always good to confirm these intuitive arguments with rigorous calculations!)

4. Suppose $X_1,\ldots,X_n$ are independent exponentially distributed random variables each with mean $1/r$. By definition, that means random variable $Y = X_1+\ldots+X_n$ is Gamma distributed with rate r and shape n. Calculate $E(Y)$ using properties of expectation to confirm that the mean of such a gamma distribution is n/r.

Answer:

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{r} = \frac{n}{r}$$

5. Suppose the sample space for r.v. X is the unit interval [0,1], and X has density function $f(x) = 4x^3$. Find the cdf of X. Then use it to calculate $P(X \in [1/2, 1])$.

   Answer:

$$P(X \in [1/2, 1]) = \int_{1/2}^{1} f(x)\,dx = \int_{1/2}^{1} 4x^3\,dx = (1)^4 - (1/2)^4 = \frac{15}{16}$$

6. Compute the coefficient of variation (CV) for a Normal random variable with mean $\mu = 18$ and variance $\sigma^2 = 9$.

   Answer:

$$CV = \frac{\sigma}{\mu} = \frac{3}{18} = \frac{1}{6} \approx 16.67\%$$

7. What is the difference between the Strong Law of Large Numbers, the Weak Law of Large Numbers, and the Central Limit Theorem?

   Answer: See definitions. The Strong LLN and Weak LLN differ in whether the result is stated in terms of almost sure convergence (SLLN) or in convergence in probability (WLLN). The former is considered "strong" because almost sure convergence implies convergence in probability (i.e., the SLLN implies the WLLN). The CLT, however, tells us even more information. Namely, the CLT tells us about the spread of the mean of a random sample, which is approximately described by a Normal distrubution.

8. Suppose continous r.v.s X_1, X_2, $\ldots$, X_k are *iid* with density function $f(x)$. Suppose $\mathbf{X}$ is the random vector $(X_1, \ldots, X_k)$. What is the joint density function of $\mathbf{X}$?

   Answer: Because they are independent, the joint pdf is

$$f_{\mathbf{X}}(x_1, x_2, ..., x_k) = \prod_{i=1}^{n} f(x_i)$$

9. Suppose there are 15 individuals in a population of 100 who carry a disease. Suppose you randomly select 50 of them, and test them for the disease, and let r.v. X be the number that actually have it (suppose a 100% accurate test). What named distribution describes the distribution of X?

   Answer: The *Binomial* distribution is a great approximation to the correct answer, however it's technically incorrect! Here, the outcome of the tests for each individual are NOT independent, since we are effectively sampling them *without replacement.* For example, if 15 of the first 49 test positive, you are 100% certain the test of the 50th will be negative. Thus, the correct distribution is the Hypergeometric distribution.