STATISTICS 571 Discussion #1

TA: Jingjiang peng
Office: 1275 MSC
Phone: 262-1577
E-mail: peng@stat.wisc.edu
Office Hours: 2:00 - 3:00pm Thursday
1:00 - 2:00pm Friday

RULES

I. Review

1. Binomial Distribution:

   (a) 'BINS'(Binary, Independent,N fixed, Same probability)

      i. each of the trial has only two outcomes, whcih are denoted as 'success' or 'failure'

      ii. a sample of n independent trials

      iii. n, the sample size is fixed

      iv. each trail has the same probability of 'success' p

   (b) Probability formula: $X \sim$ Binomial(n,p)

   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \ \ k = 0, 1, \ldots, n$$

   where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ the number of ways to choose k objects from n

   (c) X is the number of 'sucess' in the n indepedent trials

   (d) Mean $E[X] = np$, and Variance $var[X] = np(1-p)$.

2. Permutaion and Combinations

   (a) Factorial: $n! = n(n-1)\ldots 2 \times 1$

(b) Permulations(order maters): The number of permutations to take k objects from n,

$$n(n-1)\dots(n-k+1) = \frac{n!}{k!}$$

(c) Combinations(order does not mater): The number of combinations to take k objects from n

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

3. Random Variable and Discrete Distribution:

(a) A random variable is a numeric function that assigns probabilities to different events in a sample space

(b) A random variable which can be enumerated is called discrete random variable

(c) A random variable which can not be enumerated is called continuous random variable

(d) The **Expected Value** of a discrete random variable X is:

$$E[X] = \mu = \sum_{i=1}^{N} x_i P(X = x_i)$$

$x_i's$ are the values the random variable assumes with positive probability

(e) The **Variance** of a discrete random variable X is:

$$var[X] = \sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 P(X = x_i) = E[X^2] - [EX]^2$$

II.Practice Problems

1. Calcuate $\binom{10}{4}$, $\binom{25}{3}$
2. For each part, determine if the random variables is binomial or not. If so, state the values of the parameters n and p. If not, explain which assumption(s) of the binomial distribution are violated
   (a) Many couples pannning a new family would prefer to have at least one child of each sex. The probability that a couple's child is a boy is 0.512. Suppose that the sex of child is independent of the sex of former children. A new family plans to have 5 children. Let X_1 be the number of boys
   (b) Assume that the child's sex depends on the former child. For example, if the first baby is boy, then the probability that the second child is boy will decrease. A new family plans to have 5 children. Let X_2 be the number of boys
   (c) There are four blood types, A, B, AB and O. Assume that they have the same probability. We randomly select 100 people. Let X_3 be the number of type A, X_4 be the number of type B, and X_5 be the number of type AB
   (d) For these 4 types of blood, if we are just interested in type O or not type O. Let X_6 be the number of type O
3. Suppose $X \sim$ Binomal(10,0.4), calculate the followings
   (a) P(X = 3)
   (b) $P(X <= 3)$
   (c) $P(X > 4)$
   (d) $P(1 \leq X \leq 3)$, and $P(1 \leq X < 3)$
   (e) Calcuate mean and standard deviation of X
   (f) $P(E(X) - SD(X) \leq X \leq E(X) + SD(X))$
   (g) $P(E(X) - 2SD(X) \leq X \leq E(X) + 2SD(X))$

4. Many new drugs have been introduced in the last several decades to bring hypertension under control. Supppose a physician agrees to use a new antihypertensive drug on a trail basis on the first 4 untreated hypertensives she encounters in her practice, before deciding whether to adopt the drug for routine use. Let X = the number of patients out of 4 who are brought under control.Then X is a discrete random variable which takes on the value of 0, 1, 2, 3, 4. Suppose from previous experience with the drug, the drug company expects that for any clinical practice the probability that 0 patients out of 4 will be brought under control is 0.008, 1 patient out of 4 is 0.076, 2 patient out of 4 is 0.265, 3 patients out of 4 is 0.411 and all 4 patients is 0.240.

   (a) Draw the distribution table

   (b) Check that it is a correct distribution

   (c) calcuate E(X), Var(X) and SD(X)

III Solutions of the Practice problems

1. (a) $\binom{10}{4} = \frac{10\times9\times8\times7}{4\times3\times2\times1} = 210$

   (b) $\binom{25}{3} = \frac{25*24*23}{3*2} = 2300$

2. (a) Yes. $n = 5$, $p = 0.512$

   (b) No. violation of independent and constant p

   (c) No. Violation of binary outcome. Actually they are multinomial distribution

   (d) Yes. $n = 100$, $p = 0.25$

3. (a) $P(X = 3) = \binom{10}{3}0.4^3 0.6^7 = 120 * 0.064 * 0.028 = 0.215$

   (b) $P(X = 0) = \binom{10}{0}0.4^0 0.6^{10} = 1 * 1 * 0.006 = 0.006$, $P(X = 1) = \binom{10}{1}0.4^1 0.6^9 = 0.0403$, $P(X = 2) = \binom{10}{2}0.4^2 0.6^8 = 0.1209$

$$
\begin{aligned}
P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
&= 0.006 + 0.0403 + 0.1209 + 0.215 = 0.3822
\end{aligned}
$$

   (c) $P(X > 4) = 1 - P(X <= 3) = 1 - 0.3822 = 0.6178$

   (d)

$$P(1 \leq X \leq 3) = P(X = 1)+P(X = 2)+P(X = 3) = 0.0403+0.1209+0.215 = 0.3762$$

$$P(1 \leq X < 3) = P(X = 1) + P(X = 2) = 0.0403 + 0.1209 = 0.1612$$

   (e) $EX = np = 4, var(X) = np(1 - p) = 2.4, SD(X) = \sqrt{var(X)} = 1.55$

   (f)

$$
\begin{aligned}
P(4 - 1.55 \leq X \leq 4 + 1.55) &= P(2.45 \leq X \leq 5.55) \\
&= P(X = 3) + P(X = 4) + P(X = 5) \\
&= 0.215 + 0.2508 + 0.2006 = 0.6664
\end{aligned}
$$

(g)

$$P(4-2*1.55 \le X \le 4+2*1.55) = P(0.9 \le X \le 7.1)$$
$$= 1 - P(X=0) - P(X=8) - P(X=9) - P(X=10)$$
$$= 1 - 0.006 - 0.0106 - 0.0016 - 0.0001 = 0.9817$$

4. (a)

k	0	1	2	3	4
P(X=k)	0.008	0.078	0.265	0.411	0.240

(b) all probability is positive and 0.008+0.078+0.265+0.411+0.240 = 1

(c)

$$E(X) = 0*0.008 + 1*0.076 + 2*0.265 + 3*0.411 + 4*0.240 = 2.80$$
$$E(X^2) = 0^2*0.008 + 1^2*0.076 + 2^2*0.265 + 3^2*0.411 + 4^2*0.240 = 8.675$$
$$Var(X) = E(X^2) - (EX)^2 = 8.675 - 2.80^2 = 0.835$$
$$SD(X) = \sqrt{Var(X)} = 0.9138$$