MAST90105: Lab and Workshop 4

1 Lab

In this lab, we will examine the shape of distributions for different values of their parameters.

1. Download and open the file lab4.R in R or RStudio

   a. The function `m34binomial(n,p)` returns a skewness (first column) and kurtosis (second column) of a binomial distribution with parameters n and p (these can be vectors of the same length). Here, n is the number of trials and p is the probability of success in each trial. Use this function to compute the skewness and kurtosis and record these values in the following table for the **binomial distribution**:

	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
n=1					
n=10					
n=50					
n=100					
n=250					
n=500					

   b. Comment on the trends you find in your table, in particular across the rows and down the columns

   c. The function `m34negbinomial(r,p)` returns a skewness (first column) and kurtosis (second column) of a negative binomial distribution with parameters r and p (these can be vectors of the same length). Here, experiment continues until r successes and p is the probability of success in each trial. Use this function to compute the skewness and kurtosis and record these values in the following table for the **negative binomial distribution**:

	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
r=1					
r=10					
r=50					
r=100					
r=250					
r=500					

d. Comment on the trends you find in the table, in particular across the rows and down the columns

2. The function `distBinomHyper(n,p,t)` computes the distance between the binomial distribution Binom(n, p) and the hypergeometric distribution Hyper$(b, t-b, n)$ with $b = p \cdot t$. This function also plots these two pmf-s for all values from the interval $\left(np \pm 5\sqrt{np(1-p)}\right)$. Here, n is the sample size, p is the probability of success and the proportion of items of type 1 for the binomial and hypergeometric distributions, respectively, and b, $t-b$ are the numbers of items of type 1 and 2, respectively, in the population of size t. The distance between the two distributions is defined as

$$D = \max_A |\Pr(X \in A) - \Pr(Y \in A)| = 0.5 \sum_i |\Pr(X = i) - \Pr(Y = i)|\,. \tag{1}$$

a. Take the sample size n to be 500, 1000 and 3000. Record the three distances between the binomial and hypergeometric distributions for each of the combinations in the table:

	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
t=n+100					
t=n+500					
t=n+1,500					
t=n+2,000					
t=n+10,000					
t=16,000,000					

b. Comment on the trends in the table, in particular across the rows and columns

c. Looking at the R code, what is the purpose of defining the variable `i.plot`?

d. (*Challenging!*) Prove the equality (1) for any discrete random variables X and Y. (Hint: use the fact that probabilities add to one, and think about the set of numbers for which the hypergeomtric probability is greater than the binomial probability and vice-versa.)

3. The objective of this question is to use R to look at the actual distance between binomial and Poisson distributions.

   a. We learned that the maximum difference between any binomial and the corresponding Poisson probability, denoted D in the plots, was bounded by p, the probability of success in the binomial distribution (check module 2, Section 7.1). Modify the function `distBinomHyper(n,p,t)` to compute the distance between the Binomial distribution Binom(n, p) and the Poisson distribution Poisson(λ), with $\lambda = np$, and to plot their pmf-s. Choosing a variety of values for p between 0 and 1, compute and record the values of D, for different values of n. Comment, particularly on whether the relationship with n is monotone.

   b. Find the maximum difference $p - D$ and write it down. What do the plots in (b) show about the guidance that is given in many textbooks that "the Poisson approximation to Binomial can be used when n is large, p is small and np is moderate"?

   c. Try to find n and p that minimize the ratio between p/D. Report on your findings.

2 Workshop

geometric distribution 最后一次损坏一次

1. A warranty is written on a product worth \$10,000 so that the buyer is given \$8000 if it fails in the first year, \$6000 if it fails in the second, \$4000 if it fails in the third,

\$2000 if it fails in the fourth, and zero after that. Its probability of failing in a year is 0.1; failures are independent of those of other years. What is the expected value of the warranty?

2. Define the pmf and give the values of μ and σ^2 when the moment-generating function (mgf) of X is defined by

   a. $M(t) = \left(\frac{0.6e^t}{1-0.4e^t}\right)^2, \quad t < -\ln(0.4).$

   negative Binomial $\binom{x}{2}(0.6)^2(0.4)^{x-2}$ $\mu = \frac{1}{p} = \frac{2}{0.6}$ $Var = \frac{1-p}{p^2} = \frac{0.4\times 2}{0.6^2}$

3. Let X equal the number of people selected at random that you must ask in order to find someone with the same birthday as yours. Assuming each day of the year is equally likely (and ignoring February 29).

   geometric $P(X=k) = \left(\frac{1}{365}\right)\left(1-\frac{1}{365}\right)^{k-1}$

   a. What probability distribution does X have? Namely, what is the pmf of X?

   b. Give the mean and variance of X.

   mean $= \frac{1}{p} = 365$ $Var = \frac{1-p}{p^2} = \frac{\frac{364}{365}}{\left(\frac{1}{365}\right)^2}$

   c. Find $P(X > 400)$ and $P(X < 300)$.

   $P(X>400)$ 前400都失败 $\left(1-\frac{1}{365}\right)^{400}$

   $P(X<300) = 1-P(X\geq 300) = 1-P(X=300)-P(X>300) = 1-\left(1-\frac{1}{365}\right)^{299}\left(\frac{1}{365}\right)-\left(1-\frac{1}{365}\right)^{300}$

4. Let X equal the number of flips of a fair coin that are required to observe the same face on consecutive flips. Find the pmf of X.

5. Suppose that a basketball player can make a free throw 60% of the time. Let X equal the minimum number of free throws that this player must attempt to make a total of 10 shots,

   →罚球

   Nega B $\binom{x-1}{9}(0.6)^9\times 0.4^{(x-1-9)}\times 0.6$

   a. What probability distribution does X have? Namely, what is the pmf of X?

   b. Give the mean and variance of X.

   $\frac{r}{p} = \frac{10}{0.6}$ $\frac{10(1-0.6)}{(0.6)^2}$

   c. Find $P(X = 16)$.

6. Let X have a Poisson distribution with a variance of 3. Find $P(X = 2)$.

7. Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume the Poisson distribution, find the probability of at most one flaw in 225 square feet.

   $\lambda = \left(\frac{1}{150}\times\frac{150}{225}\right)$ $\left(\frac{e^{-\frac{1}{225}}\left(\frac{1}{225}\right)^0}{0!}\right) + \left(\frac{e^{-\frac{1}{225}}\times\frac{1}{225}}{1!}\right) = \frac{226}{225}e^{-\frac{1}{225}}$ $\frac{150}{225}$

8. Suppose that the probability of suffering a side effect from a certain flu vaccine is 0.005. Also suppose 1000 persons are inoculated.

   a. Find the exact probability that at most 1 person suffers using a binomial distribution.

   $\binom{1}{1000}(5\times 10^{-3}) + (0.995)^{1000}$

   b. Find approximately the probability that at most 1 person suffers using a Poisson distribution.

   $np = \lambda$

9. A hospital obtains 40% of its flu vaccine from Company A, 50% from Company B, and 10% from Company C. From past experience it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective. The hospital test 5 vials from each shipment. If at least one of the five is ineffective, find the conditional probability of that shipment coming from C.

   $P_A = 1-(1-0.03)^5$ $P_B = 1-(1-0.02)^5$ $P_C = 1-(1-0.05)^5$

   P_n(至少有一个ineffective) $= (1-0.03)^5\times\frac{5}{0.4x} + (1-0.02)^5\times\frac{5}{0.5x} + (1-0.05)^5\times\frac{5}{0.1x} \approx 0.524$

   P(来自C | 至少一个坏) P(至少一个坏 | 来自C) P(来自C) $(1-0.05)^5\times\frac{5}{0.1}$

Pr(第4套试卷) = $\frac{\ldots}{0.524}$ = $\frac{\ldots}{0.524}$ = $\boxed{0.97^5 \times \frac{5}{0.14} + 0.98 \times \frac{5}{0.5} + 0.95 \times \frac{5}{0.1}}$

10. If X has a Poisson distribution so that $3P(X = 1) = P(X = 2)$, find $P(X = 4)$.

11. One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased?

4 X be ~~number~~ number of flips of a fair coin that are required to observe same face on consequtive flips (PMF of X)

2 3 4 5

THH HTT

OO $\frac{1}{2} \times \frac{1}{2}$ OOO $\frac{1}{2} \times (\frac{1}{2})^2$ OOOO $(\frac{1}{2}) \times (\frac{1}{2})^2$ OOOOO $(\frac{1}{2})^3$

$\boxed{2x \left(\frac{1}{2}\right)^{x-2} \left(\frac{1}{2}\right)^2}$

H

$\frac{1}{2}\left(\frac{1}{2}\right)^{k-2}$

HTHH → $\frac{1}{2} \times \frac{1}{2} \times (\frac{1}{2})^2$

THTT $\frac{1}{2} \times \frac{1}{2} \times (\frac{1}{2})^2$

$\frac{1}{2} \times \frac{1}{2}$

6 HTHTHH

THTHTT

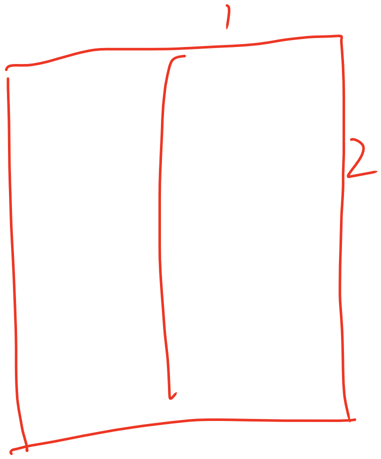
$(\frac{1}{2})^4$ $\frac{1}{2} \times \frac{1}{2}$

Check. $\sum_{k=2}^{\infty} \frac{1}{2} \times \left(\frac{1}{2}\right)^{k-2} = \frac{1}{2} \sum_{k=2}^{\infty} \left(\frac{1}{2}\right)^{k-2}$ $i+2=k$

$= \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i$

$= \frac{1}{2} \times 2 = 1$

1 切正方形面积



考试时：☆

$\{X=3\} = \{HTT\} \cup \{THH\}$

$\{X=5\} = \{HTHTT\} \cup \{THTHH\}$

PMF of X $P(X=k)$, $k=2,3,4,5\dots\dots$

每个盒子都有prize.

11. One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased?

multiple geometric distribution

total : n 4 A B C D

$\binom{n-1}{}(p)^3 \quad (\ )^1$

your r

T_1: # boxes until get 2 unique prizes after 1st box ~ geo

$P(T_1=2) = (\frac{1}{4})^{2-1}(\frac{3}{4}) \qquad E(T_1) = \frac{1}{p} = \frac{4}{3}$

T_2: # boxes after T_1 until get 3 unique prizes

$\ \ ) = (\frac{2}{4})^{2-1}(\frac{2}{4}) \qquad E(T_2) = \frac{1}{p} = \frac{4}{2} = 2$

$P(T_2=2)=$

T_3: # after T_2 until get you 4 unique prizes

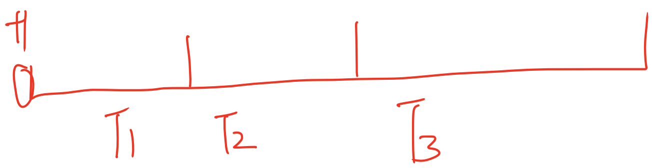
$E(T_3)=\frac{1}{p}=\frac{4}{1}=4$

$P(T_3=i)=\left(\frac{3}{4}\right)^{i-1}\left(\frac{1}{4}\right)$

Ans: $E(T_1+T_2+T_3)$

#

$= 1 + E(T_1)+E(T_2)+E(T_3) = 1+\frac{4}{3}+2+4$



geometric distribution

1. A warranty is written on a product worth \$10,000 so that the buyer is given \$8000 if it fails in the first year, \$6000 if it fails in the second, \$4000 if it fails in the third,

\$2000 if it fails in the fourth, and zero after that. Its probability of failing in a year is 0.1; failures are independent of those of other years. What is the expected value of the warranty?

X	8000	6000	4000	2000	
P	0.1	0.1×0.9	$(0.1)\times0.9^2$	0.1×0.9^3	

$800 + 60\times9 + 4\times81 + 2\times0.9\times81$

$= 1809.8$

2. Define the pmf and give the values of μ and σ^2 when the moment-generating function (mgf) of X is defined by

negative Binomial $\binom{x}{2}(0.6)^2(0.4)^{x-2}$ Var $= \frac{1-p}{p^2} =$

$M = \frac{1}{p} = \frac{2}{0.6}$

a. $M(t) = \left(\frac{0.6e^t}{1-0.4e^t}\right)^2, \quad t < -\ln(0.4).$

$\frac{2}{0.6}$

3. Let X equal the number of people selected at random that you must ask in order to find someone with the same birthday as yours. Assuming each day of the year is equally likely (and ignoring February 29).

geometric $P(X=k) = \left(\frac{1}{365}\right)\left(1-\frac{1}{365}\right)^{k-1}$

a. What probability distribution does X have? Namely, what is the pmf of X?

b. Give the mean and variance of X.

c. Find $P(X > 400)$ and $P(X < 300)$.

mean $= \frac{1}{p} = 365$ var $= \frac{1-p}{p^2} = \frac{\frac{364}{365}}{\left(\frac{1}{365}\right)^2}$

$\frac{1}{365}$ $\frac{364}{365}$ geometric $1-P(X>800)$

$\frac{\frac{1}{365}e^t}{1-\frac{1}{365}}$

$P(X>400) = 1-P(X\le 400)$

$= 1-\sum_{x=1}^{400} \frac{1}{365} \times \left(\frac{364}{365}\right)^{x-1}$

$P(X<300)$ $1-P(X\ge 300)$

4. Let X equal the number of flips of a fair coin that are required to observe the same face on consecutive flips. Find the pmf of X.

$X=2$ HH、TT $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^2 \times 2$

$X=3$ HTT、THH $\frac{1}{2}\times\left(\frac{1}{2}\right)^2 + \frac{1}{2}\times\left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^3 \times 2$

$X=4$ HTHH, THTT $\frac{1}{2}\times\frac{1}{2}\times\left(\frac{1}{2}\right)^2 + \frac{1}{2}\times\frac{1}{2}\times\left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^4 \times 2$

$X=n$: $\left(\frac{1}{2}\right)^n \times 2 = \frac{1}{2^{n-1}}$ $k=2.3.4.5$

5. Suppose that a basketball player can make a free throw 60% of the time. Let X equal the minimum number of free throws that this player must attempt to make a total of 10 shots,

Nega B $\binom{x-1}{9}(0.6)^9 \times 0.4^{x-1-9} \times 0.6$

   a. What probability distribution does X have? Namely, what is the pmf of X?
   b. Give the mean and variance of X. $\frac{r}{p} = \frac{10}{0.6}$ $\frac{10(1-0.6)}{(0.6)^2}$ 150
   c. Find $P(X = 16)$.

$P = 0.6$ # 10 shots made

$X \sim NB(10, 0.6, 0.4)$

$$\frac{(0.6e^t)^{10}}{(1-0.4e^t)^{10}}$$

$mean = \frac{r}{p} = \frac{100}{0.6} = \frac{50}{3}$

$\boxed{10 \times \frac{0.4}{0.6^2}} = \frac{40 \times 10}{6 \times 6}$

$\binom{16}{10} 0.6^{10} 0.4^{6}$

$\lambda = 3$

$\lambda = \frac{1}{150} \to 225$

$\lambda = \frac{225}{150}$

$$\sum_{i=0}^{1} \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\frac{225}{150}} + e^{-\frac{225}{150}} \times \frac{225}{150}$$

1%的 $P = \lambda 0.005$ 最多一个 0, 1 0.44

100人接打 1

$$\binom{1000}{1}(0.005)^1(0.995)^{999} + (0.995)^{1000}$$

$np = \lambda$

$\lambda = 1000 \times 0.005 = 5$

$$\sum_{i=0}^{1} \frac{e^{-5} 5^i}{i!} = e^{-5} + 5e^{-5} = 6e^{-5}$$

LOTP:

P(ineffective) $4\% \times 3\% + 5\% \times 2\% + 1\% \times 5\% = 0.4 \times 0.03 + 0.5 \times 0.02 + 0.1 \times 0.05$

$= 0.005 + 0.010 + 0.012$

$$P(C \mid \text{ineffect}) = \frac{P(C, \text{ineffective})}{P(\text{ineffective})} = \frac{0.005}{0.027} = 0.022 + 0.005 = 0.027$$

$= \frac{5}{27}$

11. One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased?

every prizes is a geometric distribution

11. One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased?

$1 + \left(\frac{3}{4}\right) + \frac{2}{4} + \frac{1}{4} = \frac{4}{4} + \frac{6}{4} = \frac{10}{4} = \frac{5}{2}$

$1 + \frac{4}{3} + \frac{4}{2} + 4 = 5 + 2 + \frac{4}{3}$

$= 8\frac{1}{3}$

$1 + X_1$

$r = \frac{3}{4}$ Geometric