

MAST90105 Lab and Workshop 3 Solutions

1 Lab

In this lab., you may use either R or Mathematica or both. Some questions may be easier in one or the other.

1. An urn contains 17 balls marked LOSE and 3 balls marked WIN. You and an opponent take turns selecting at random a single ball from the urn without replacement. The person who selects the third WIN ball wins the game. It does not matter who selected the first two WIN balls.

- a. If you draw first, find the probability that you win the game on your second draw.

Solution:

You win on your second draw if both you and your opponent choose WINS on the first two draws and you then choose a WIN on the third draw as well. This is the same as drawing three WINs in a random sample of size 3 from the bag. Hence

$$P(\text{You win on your second draw}) = \frac{\binom{3}{3} \binom{17}{0}}{\binom{20}{3}} = \frac{1}{1140}.$$

You can find this from Mathematica as follows:

PDF[HypergeometricDistribution[3, 3, 20], 3]

$$\frac{1}{1140}$$

N[%]

0.000877193

An alternative way to think about this probability uses the multiplication rule to express the probability as the product of the probability of getting 2 WINS in the first two draws with the conditional probability of getting a WIN on the third draw given you got a WIN on both of the first two draws. Hence,

$$P(\text{You win on your second draw}) = \frac{\binom{2}{2} \binom{17}{0}}{\binom{20}{2}} \times \frac{1}{18} = \frac{1}{1140}.$$

You can find this from Mathematica as follows and the fractional answer is identical with the previous calculation.

PDF[HypergeometricDistribution[2, 3, 20], 2] * 1/18

$$\frac{1}{1140}$$

The same calculations can be done in R as follows:

```
dhyper(x = 3, m = 3, n = 17, k = 3)
## [1] 0.000877193

dhyper(x = 2, m = 3, n = 17, k = 2) * 1/18
## [1] 0.000877193
```

- b. If you draw first, find the probability that your opponent wins the game on his second draw.

Solution:

Your opponent wins on their second draw, if there are two WINS in the first three draws from the urn and the fourth draw, which is the second one for the opponent, is the last WIN. Using similar reason as the alternative in part (a),

$$P(\text{Your opponent wins on his second draw}) = \frac{\binom{3}{2}\binom{17}{1}}{\binom{20}{3}} \times \frac{1}{17} = \frac{1}{380}.$$

This can be found in Mathematica by:

PDF[HypergeometricDistribution[3, 3, 20], 2] * 1/17

$$\frac{1}{380}$$

N[%]

0.00263158

The same calculations can be done in R as follows:

```
dhyper(x = 2, m = 3, n = 17, k = 3) * 1/17
## [1] 0.002631579
```

- c. If you draw first, the probability that you win can be found from

$$P(\text{You win if you draw first}) = \sum_{k=1}^9 \frac{\binom{3}{2}\binom{17}{2k-2}}{\binom{20}{2k}} \times \frac{1}{20-2k} \quad (\text{Why?})$$

Note: You could win on your second, third, fourth, ..., or tenth draw, not on your first.

Solution:

The events that you win on your second, third, ..., tenth turn are mutually exclusive and have union the event that you win. Hence the probability that you win is the sum of the probabilities of these events. The probability you win on your

$k + 1$ th ($k = 1, \dots, 9$) turn is the k th term in the sum using the alternative reasoning in part (a): the hypergeometric probability is the probability of the event, A , of 2 WIN and $2k - 2$ LOSE balls in the first $2k$ draws from the urn, and $\frac{1}{20-2k}$ is the probability of you getting WIN on your $k + 1$ th turn conditional on the event A .

This can be found in Mathematica by:

Sum[PDF[HypergeometricDistribution[2k, 3, 20], 2] * 1/(20 - 2k), {k, 9}]

$\frac{35}{76}$

N[%]

0.460526

The same calculation can be done in R as follows:

```
(sum(dhyper(x = 2, m = 3, n = 17, k = 2 * 1:9) * 1/(20 -
  2 * 1:9)))
## [1] 0.4605263
```

- d. If you draw second, the probability that you win can be found from

$$P(\text{You win if you draw second}) = \sum_{k=1}^9 \frac{\binom{3}{2} \binom{17}{2k-1}}{\binom{20}{2k+1}} \times \frac{1}{19-2k}. \quad (\text{Why?})$$

Solution:

Using the same reasoning as (c), if you start second, the probability you win on your $k + 1$ th turn is the k th term in the sum, since the hypergeometric probability is the probability of the event, A , of 2 WIN and $2k - 1$ LOSE balls in the first $2k + 1$ draws from the urn, and $\frac{1}{19-2k}$ is the probability of you getting WIN on your $k + 1$ th turn conditional on the event A .

This can be found in Mathematica by:

Sum[PDF[HypergeometricDistribution[2k + 1, 3, 20], 2] * 1/(19 - 2k), {k, 9}]

$\frac{41}{76}$

N[%]

0.539474

The same calculation can be done in R as follows:

```
(sum(dhyper(x = 2, m = 3, n = 17, k = 2 * 1:9 + 1) *  
      1/(19 - 2 * 1:9)))  
## [1] 0.5394737
```

- e. Based on your results in (c) and (d), would you prefer to draw first or second? Why?

Solution:

Draw second, as the probability of winning is larger. Check: the probabilities have to add to one because the probability that you win if you start second is the same as the probability that your opponent wins if you start first. Either you or your opponent must win if you start first.

2. Suppose in a lot of 100 fuses there are 20 defective ones. A sample of 5 fuses are randomly selected from the lot without replacement. Let X be the number of defective fuses found in the sample.

- a. Find the probability $P(X = 0)$, *Solution:*

This can be found in Mathematica by:

PDF[HypergeometricDistribution[5, 20, 100], 0]

$\frac{19513}{61110}$

N[%]

0.319309

The same calculation can be done in R as follows:

```
dhyper(x = 0, m = 20, n = 80, k = 5)  
## [1] 0.3193094
```

- b. The cumulative probability $P(X \leq 3)$,

Solution:

This can be found in Mathematica by:

CDF[HypergeometricDistribution[5, 20, 100], 3]

$\frac{780046}{784245}$

N[%]

0.994646

The same calculation can be done in R as follows:

```
phyper(q = 3, m = 20, n = 80, k = 5)
## [1] 0.9946458
```

- c. The mean or expectation of X , $E(X)$,

Solution:

This can be found in Mathematica by:

mu = Mean[HypergeometricDistribution[5, 20, 100]]

1

or, using escape-keydistescape-key for the symbol \approx ,

Expectation[X, X \approx HypergeometricDistribution[5, 20, 100]]

1

The same calculation can be done in R as follows:

```
(mu = sum(1:5 * dhyper(x = 1:5, m = 20, n = 80, k = 5)))
## [1] 1
```

- d. The second moment of X , $E(X^2)$,

Solution:

This can be found in Mathematica by:

ex2 = Expectation[X^2, X \approx HypergeometricDistribution[5, 20, 100]]

$\frac{175}{99}$

N[%]

1.76768

The same calculation can be done in R as follows:

```
(ex2 = sum((1:5)^2 * dhyper(x = 1:5, m = 20, n = 80,
                             k = 5)))
## [1] 1.767677
```

- e. The variance of X , $\text{Var}(X)$,

Solution:

This can be found in Mathematica by:

ex2 - mu^2

$\frac{76}{99}$

or

Variance[HypergeometricDistribution[5, 20, 100]]

$\frac{76}{99}$

N[%]

0.767677

The same calculation can be done in R as follows:

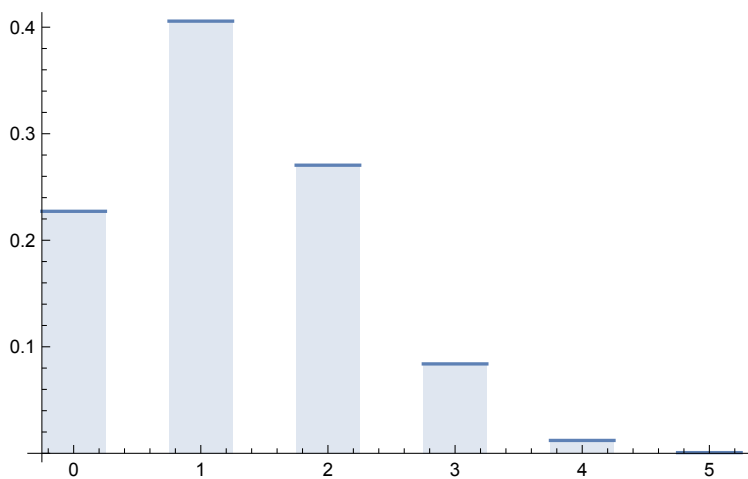
```
ex2 - mu^2
## [1] 0.7676768
```

- f. A probability bargraph for the pmf of X .

Solution:

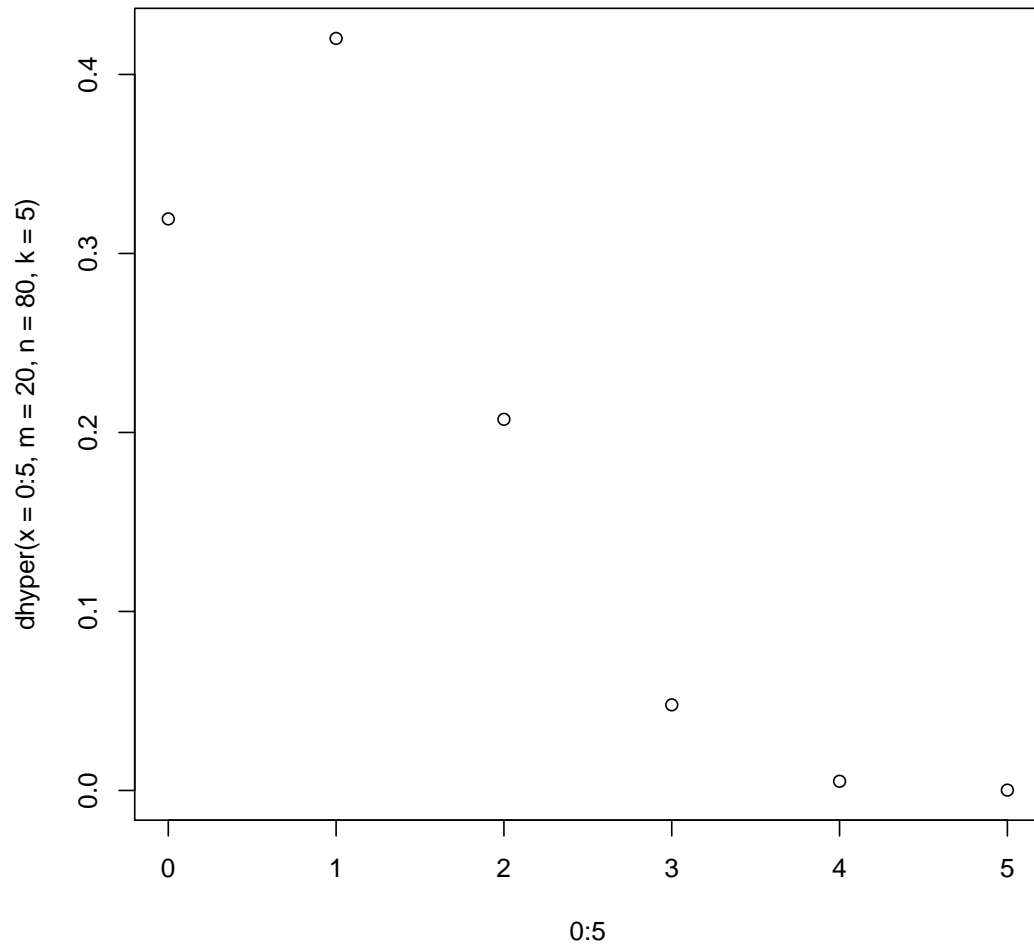
This can be found in Mathematica by:

DiscretePlot[PDF[HypergeometricDistribution[5, 20, 100], k], {k, 0, 5}, ExtentSize → 0.5]



The same plot can be done in R as follows:

```
plot(0:5, dhyper(x = 0:5, m = 20, n = 80, k = 5))
```



3. An urn contains n balls numbered from 1 to n . A random sample of n balls is selected from the urn, one at a time. A match occurs if ball numbered i is selected on the i th draw.

- If the draws are done *with replacement*, it can be shown that

$$P(\text{at least one match}) = 1 - \left(1 - \frac{1}{n}\right)^n.$$

- If the draws are done *without replacement*, it can be shown that

$$P(\text{at least one match}) = 1 - \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}\right) = 1 - \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

- a. For each value of n given in the following table, find P (at least one match) and write down the results at appropriate places of the table.

Solution:

This entries in the table can be found in Mathematica by:

Note that infinty value occurs much faster in n without replacement.

$f = \text{Function}[n, N[1 - (1 - 1/n)^n]]$

$\text{Function}[n, N[1 - (1 - \frac{1}{n})^n]]$

$f[\{1, 3, 10, 15, 100, 10000, 100000\}]$

$\{1., 0.703704, 0.651322, 0.644736, 0.633968, 0.632139, 0.632122, 0.632121\}$

$N[1 - \text{Exp}[-1]]$

0.632121

$g = \text{Function}[n, N[1 - \text{Sum}[(-1)^k / \text{Factorial}[k], \{k, 0, n\}]]]$

$\text{Function}[n, N[1 - \sum_{k=0}^n \frac{(-1)^k}{k!}]]$

$g[\{1, 3, 10, 15, 100, 10000, 100000\}]$

$\{1., 0.666667, 0.632121, 0.632121, 0.632121, 0.632121, 0.632121\}$

\vdots

```
n <- c(1, 3, 10, 15, 100)
1 - (1 - 1/n)^n

## [1] 1.0000000 0.7037037 0.6513216 0.6447356
## [5] 0.6339677

1 - exp(1)^(-1)
```



```
## [1] 0.6321206

prob <- function(n) {
  1 - sum((-1)^(0:n)/factorial(0:n))
}
lapply(n, prob)

## [[1]]
## [1] 1
##
## [[2]]
## [1] 0.6666667
##
## [[3]]
## [1] 0.6321205
##
## [[4]]
## [1] 0.6321206
##
## [[5]]
## [1] 0.6321206

1 - exp(-1)

## [1] 0.6321206
```

- b. We can also use R to simulate the processes of drawing n balls with or without replacement from a set of n balls numbered from 1 to n . We can then simulate the probability of at least one match using the relevant relative frequencies.
- Create a function `match.f` in R-Studio by writing a script and executing it as follows.

```
match.f <- function(n, simsize, rep = TRUE) {
  freq = 0
  for (i in 1:simsize) {
    sam = sample(1:n, size = n, replace = rep)
    freq = freq + (sum(sam == 1:n) >= 1)
  }
  freq/simsize
}
```

- Note that `(sum(sam==1:n)>=1)` in `match.f` is for checking whether or not there is at least 1 match in `sam`.
- Simulate the drawing process 1000 times (`simsize=1000`) for each given n and `rep` (`rep=TRUE` indicates the “with replacement” procedure is used.) Execute the following and write down the results at appropriate places in the table

that follows.

Solution:

Note that everyone's simulations will be different but that the exact calculations are much the same as the exact calculations.

```
match.f(n = 1, simsize = 1000, rep = TRUE)
## [1] 1
match.f(n = 3, simsize = 1000, rep = TRUE)
## [1] 0.686
match.f(n = 10, simsize = 1000, rep = TRUE)
## [1] 0.649
match.f(n = 15, simsize = 1000, rep = TRUE)
## [1] 0.621
match.f(n = 100, simsize = 1000, rep = TRUE)
## [1] 0.619
match.f(n = 10000, simsize = 1000, rep = TRUE)
## [1] 0.613
match.f(n = 1, simsize = 1000, rep = FALSE)
## [1] 1
match.f(n = 3, simsize = 1000, rep = FALSE)
## [1] 0.666
match.f(n = 10, simsize = 1000, rep = FALSE)
## [1] 0.616
match.f(n = 15, simsize = 1000, rep = FALSE)
## [1] 0.635
match.f(n = 100, simsize = 1000, rep = FALSE)
## [1] 0.624
match.f(n = 10000, simsize = 1000, rep = FALSE)
## [1] 0.631
```

n	$P(\text{at least one match})$			
	with replacement		without replacement	
	by Calculation	by R simulation	by Calculation	by R simulation
1				
3				
10				
15				
100				
∞				

2 Workshop

- Let $f(x) = (1/4)^{|x|}(1/2)^{1-|x|}$ for $x = -1, 0, 1$ and $f(x) = 0$ for other x values. Find the mean, variance and moment generating function for a random variable with this pmf.

- It follows from the formula that $f(-1) = 1/4$, $f(0) = 1/2$, $f(1) = 1/4$ and $f(x) = 0$ for all other x values. The range of X is $R_X = \{-1, 0, 1\}$; $f(x) > 0$ for any $x \in R_X$ and the total probability $f(-1) + f(0) + f(1) = 1$. Therefore, $f(x)$ is a pmf and can be expressed by the following table:*

x	-1	0	1
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$E(X) = 0$, by symmetry. X^2 only takes the values 0 and 1, so $E(X^2) = P(X^2 = 1) = 1 - P(X = 0) = \frac{1}{2}$. Hence $\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{2}$. The moment generating function is $M(t) = e^{-t}f(-1) + e^{0t}f(0) + e^t f(1) = \frac{e^{-t}+e^t}{4} + \frac{1}{2}$.

- Let a random experiment be the cast of a pair of unbiased 6-sided dice and let X equal the smaller of the outcomes if they are different and the common value if they are equal.
 - With reasonable assumptions, find the pmf of X .
 - The pmf of X is*

x	1	2	3	4	5	6
$f(x)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

- Alternatively, the pmf of X is $f(x) = \frac{13-2x}{36}$, $x = 1, 2, 3, 4, 5, 6$.

- b. Let Y equal the range of the two outcomes (i.e., the absolute value of the difference of the largest and smallest outcomes). Determine the pmf of Y .

- The pmf of Y is

y	0	1	2	3	4	5
$g(y)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

- Alternatively, the pmf of Y is $g(y) = \left(\frac{6}{36}\right)^{1-\min(1,y)} \left(\frac{12-2y}{36}\right)^{\min(1,y)}$, $y = 0, 1, 2, 3, 4, 5$.

3. In a lot of 100 light bulbs, there are 5 bad bulbs. An inspector inspects 10 bulbs selected at random. Let X be the number of bad bulbs in the sample.

- a. What probability distribution does X have?

- X has a hypergeometric distribution $\text{Hyper}(N_1, N_2, n)$ with $N_1 = 5$, $N_2 = 95$ and $n = 10$.

- b. Calculate the probability that at least one defective bulb will be found in the sample.

•

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{5}{0} \binom{95}{10}}{\binom{100}{10}} = 1 - \frac{95 \times 94 \times \cdots \times 86}{100 \times 99 \times \cdots \times 91} = 0.416.$$

- c. Find the mean of X , i.e. $E(X)$.

- $E(X) = n \left(\frac{N_1}{N_1 + N_2} \right) = 10 \left(\frac{5}{100} \right) = 0.5$

- d. Find the variance of X , i.e. $\text{Var}(X)$. (Note that variance for Hypergeometric has not yet been discussed in class - it will be discussed in Module 4. The formula is given in the table in the textbook.)

- $\text{Var}(X) = n \left(\frac{N_1}{N} \right) \left(\frac{N_2}{N} \right) \left(\frac{N-n}{N-1} \right) = 10 \left(\frac{5}{100} \right) \left(\frac{95}{100} \right) \left(\frac{90}{99} \right) = 0.432$

- e. Find the second moment of X , i.e. $E(X^2)$.

- Since $\text{Var}(X) = E(X^2) - [E(X)]^2$,
it follows that $E(X^2) = \text{Var}(X) + [E(X)]^2 = 0.432 + 0.5^2 = 0.682$.

4. Given $E(X + 4) = 10$ and $E[(X + 4)^2] = 116$, determine

- a. $\text{Var}(X + 4)$.

- $\text{Var}(X + 4) = E[(X + 4)^2] - [E(X + 4)]^2 = 116 - 10^2 = 16$.

b. μ .

- $\mu = E(X) = E(X + 4) - 4 = 10 - 4 = 6$.

c. σ^2 .

- $116 = E[(X + 4)^2] = E(X^2 + 8X + 16) = E(X^2) + 8E(X) + 16 = \text{Var}(X) + [E(X)]^2 + 8\mu + 16 = \sigma^2 + \mu^2 + 8\mu + 16 = \sigma^2 + 6^2 + 48 + 16 = \sigma^2 + 100$. So $\sigma^2 = 16$.

5. A box contains 4 coloured balls: 2 black and 2 white. Balls are randomly drawn successively without. If X is the number of draws until the last black ball is obtained, what are the possible values of X ? Find the pmf $f(x)$ for X . (*Hint*: Define events $B_i = \{\text{the } i\text{-th draw is a black ball}\}$ and $W_j = \{\text{the } j\text{-th draw is a white ball}\}$. Then find how each outcome of X is related to B_i and W_j .)

- The possible values of X are 2, 3, and 4.
- $P(X = 2) = P(B_1 \cap B_2) = \frac{2}{4} \times \frac{1}{3} = \frac{1}{6}$.
- $P(X = 3) = P((B_1 \cap W_2 \cap B_3) \cup (W_1 \cap B_2 \cap B_3))$
 $= P(B_1 \cap W_2 \cap B_3) + P(W_1 \cap B_2 \cap B_3) = \frac{2}{4} \times \frac{2}{3} \times \frac{1}{2} + \frac{2}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{2}{6}$.
- $P(X = 4) = P((B_1 \cap W_2 \cap W_3 \cap B_4) \cup (W_1 \cap B_2 \cap W_3 \cap B_4) \cup (W_1 \cap W_2 \cap B_3 \cap B_4))$
 $= P(B_1 \cap W_2 \cap W_3 \cap B_4) + P(W_1 \cap B_2 \cap W_3 \cap B_4) + P(W_1 \cap W_2 \cap B_3 \cap B_4)$
 $= \frac{2}{4} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{1} + \frac{2}{4} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{1} + \frac{2}{4} \times \frac{1}{3} \times \frac{2}{2} \times \frac{1}{1} = \frac{3}{6}$.
- So the pmf of X is $f(x) = \frac{x-1}{6}$, $x = 2, 3, 4$.

6. Let X be the number of accidents in a factory per week having pmf

$$f(x) = \frac{1}{(x+1)(x+2)}, \quad x = 0, 1, 2, \dots$$

a. Find the conditional probability of $X \geq 4$, given that $X \geq 1$. (*Hint*: Write $f(x) = \frac{1}{x+1} - \frac{1}{x+2}$.)

• First

$$\begin{aligned} P(X \geq 4 | X \geq 1) &= \frac{P((X \geq 4) \cap (X \geq 1))}{P(X \geq 1)} = \frac{P(X \geq 4)}{1 - P(X = 0)} \\ &= \frac{1 - \{P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\}}{1 - P(X = 0)}. \end{aligned}$$

- $P(X = 0) = \frac{1}{(0+1)(0+2)} = \frac{1}{2}$.
- $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = [\frac{1}{0+1} - \frac{1}{0+2}] + [\frac{1}{1+1} - \frac{1}{1+2}] + [\frac{1}{2+1} - \frac{1}{2+2}] + [\frac{1}{3+1} - \frac{1}{3+2}] = 1 - \frac{1}{5} = \frac{4}{5}$.
- So $P(X \geq 4 | X \geq 1) = \frac{1 - \frac{4}{5}}{1 - \frac{1}{2}} = \frac{2}{5}$.

b. Does $E(X)$ exist? If yes, find it; if not, why?

- $E(X) = \sum_{x=0}^{\infty} x \cdot \frac{1}{(x+1)(x+2)} = \sum_{x=0}^{\infty} \frac{x+1-1}{(x+1)(x+2)}$
 $= \sum_{x=0}^{\infty} \frac{1}{x+2} - \sum_{x=0}^{\infty} \frac{1}{(x+1)(x+2)} = +\infty.$
- So $E(X)$ does not exist.

7. Suppose that a school has 20 classes: 16 with 25 students in each, three with 100 students in each, and one with 300 students for a total of 1000 students.

a. What is the average class size?

- $\text{averagesize} = \frac{16 \times 25 + 3 \times 100 + 1 \times 300}{20} = 50.$

b. Select a student randomly out of the 1000 students. Let the random variable X equal the size of the class to which this student belongs. Find the pmf of X .

x	25	100	300
$f(x) = P(X = x)$	$\frac{16 \times 25}{1000} = 0.4$	$\frac{3 \times 100}{1000} = 0.3$	$\frac{1 \times 300}{1000} = 0.3$

c. Find $E(X)$, $Var(X)$ and the mgf of X .

- $E(X) = 25 \times 0.4 + 100 \times 0.3 + 300 \times 0.3 = 130,$
- $E(X^2) = 25^2 \times 0.4 + 100^2 \times 0.3 + 300^2 \times 0.3 = 30250,$
- $Var(X) = E(X^2) - \{E(X)\}^2 = 30250 - 130^2 = 13350,$
- Mgf of X $M(t) = E(e^{Xt}) = e^{25t} \times 0.4 + e^{100t} \times 0.3 + e^{300t} \times 0.3.$

8. A certain type of mint has a label weight of 20.4 grams. Suppose that the probability is 0.90 that a mint weighs more than 20.7 grams. Let X equal the number of mints that weigh more than 20.7 grams in a sample of 8 mints selected at random.

a. How is X distributed if we assume independence?

- $X \stackrel{d}{=} b(8, 0.9).$

b. Find $P(X = 8)$ and $P(X \leq 7)$.

- $P(X = 8) = 0.9^8 = 0.4305.$
- $P(X \leq 7) = 1 - P(X = 8) = 0.5695.$

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

a. $M(t) = \frac{1}{3} + \frac{2}{3}e^t.$

- X has a Bernoulli distribution $b(1, 2/3).$
- Pmf $f(x) = (2/3)^x(1/3)^{1-x}, x = 0, 1.$
- $\mu = 2/3$ and $\sigma^2 = 2/9.$

b. $M(t) = (0.25 + 0.75e^t)^{12}.$

- X has a binomial distribution $b(12, 0.75).$
- Pmf $f(x) = \binom{12}{x} 0.75^x 0.25^{12-x}, x = 0, 1, 2, \dots, 12.$

- $\mu = np = 9$ and $\sigma^2 = npq = 9/4$.

10. If the moment-generating function of X is

$$M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t},$$

Find the mean, variance, and pmf of X .

- The pmf of X is

x	1	2	3
$f(x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

- $\mu = 1 \times \frac{2}{5} + 2 \times \frac{1}{5} + 3 \times \frac{2}{5} = 2$, $E(X^2) = 1^2 \times \frac{2}{5} + 2^2 \times \frac{1}{5} + 3^2 \times \frac{2}{5} = \frac{24}{5}$, and $\sigma^2 = E(X^2) - \mu^2 = \frac{24}{5} - 2^2 = \frac{4}{5}$.
- Using mgf: $\mu = M'(t)|_{t=0} = \left(e^t \times \frac{2}{5} + 2e^{2t} \times \frac{1}{5} + 3e^{3t} \times \frac{2}{5}\right) \Big|_{t=0} = 2$,
- $E(X^2) = M''(t)|_{t=0} = \left(e^t \times \frac{2}{5} + 2^2e^{2t} \times \frac{1}{5} + 3^2e^{3t} \times \frac{2}{5}\right) \Big|_{t=0} = \frac{24}{5}$.