Math 10B with Professor Stankova
Worksheet, Discussion #13; Thursday, 3/1/2018
GSI name: Roy Zhao

1 Probability Distributions

1.1 Concepts

1.

Distribution	PMF	Example
Uniform	If $\#R(X) = n$, then $f(x) = \frac{1}{n}$ for all $x \in R(X)$.	Dice roll, $f(1) = f(2) = \cdots = f(6) = \frac{1}{6}$.
Bernoulli Trial	$f(0) = 1 - p$, $f(1) = p$	Flipping a biased coin
Binomial	$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$	p is probability of success. Repeat n Bernoulli trials. Number of 6's rolled when rolling 10 die is $f(k) = \binom{10}{k}(1/6)^k(5/6)^{10-k}$.
Geometric	$f(k) = (1-p)^k p$	k failures and then a success.
Hyper-Geometric	$f(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	Counting the number of red balls I pick out of n balls drawn if there are m red balls out of N balls total.
Poisson	$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	Count the number of babies born today if on average there are 3 babies born a day.

1.2 Examples

2. On average, there are 20 rainy days in Berkeley per year. What is the probability that this year, there are 30?

> **Solution:** This is a Poisson distribution since we are talking about the average number of times something occurs in a place or over some period of time. The average is $\lambda = 20$ and hence $P(X = 30) = f(30) = \frac{20^{30}e^{-20}}{30!}$.

3. The probability of seeing a shiny Pokemon is approximately 1 in $10000 = 10^4$. What is the probability that I don't see any in my playthrough if I see 10^5 Pokemon total? (calculate both exactly and an approximation)

Solution: Assuming each Pokemon sighting is independent, this is binomial process with $p = 1/10^4$. We see 10^5 Pokemon and want to see $k = 0$ shiny Pokemon so the answer is $\binom{10^5}{0} p^0 (1-p)^{10^5} = (1 - \frac{1}{10^4})^{10^5}$.

Now a Poisson distribution approximates a binomial one whenever n is very large and p is very small. The average number of Pokemon we can expect to see when seeing 10^5 Pokemon is $np = \lambda = 10$ shiny Pokemon. Thus, the probability that we see 0 is $f(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-10}$.

1.3 Problems

4. When a cell undergoes mitosis, the number of mutations that occurs is Poisson distributed and an average of 11 mutations occur. What is the probability that no more than 1 mutation occurs when a cell divides?

Solution: This is a Poisson process with $\lambda = 11$. We want to know $P(X \leq 1) = P(X = 0) + P(X = 1) = f(0) + f(1) = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} = e^{-11} + 11e^{-11} = 12e^{-11}$.

5. The number of chocolate chips in a cookie is Poisson distributed with an average of 15 chocolate chips. What is the probability that you pick up a cookie with only 10 chocolate chips in it?

Solution: This is a Poisson distribution with $\lambda = 15$. We want to calculate $f(10) = \frac{\lambda^{10} e^{-\lambda}}{10!} = \frac{15^{10} e^{-15}}{10!}$.

6. The number of errors on a page is Poisson distributed with approximately 1 error per 100 pages of a book. What is the probability that a novel of 300 pages contains no errors?

Solution: This is a Poisson distribution and the probability that we have 300 pages with no errors means that the first 100 pages has no errors, the second 100 and the last 100 all have no errors. So the answer is $f(0)^3 = e^{-3\lambda} = e^{-3}$.

Another way to do this is noting that if we average 1 error per 100 pages, then over a novel of 300 pages, we should expect $300/100 \cdot 1 = 3$ errors. Thus, the probability of having no errors with $\lambda = 3$ is $e^{-\lambda} = e^{-3}$.

7. Approximately 4 people are born every second. What is the probability that in a minute, there are 240 people born?

Solution: Since approximately 4 people are born every second and there are 60 seconds in a minute, approximately $4 \cdot 60 = 240$ people are born every minute. Thus, this is Poisson distribution with $\lambda = 240$ and we want to calculate $f(240) = \frac{\lambda^{240} e^{-\lambda}}{240!} = \frac{240^{240} e^{-240}}{240!}$.

1.4 Extra Problems

8. When a cell undergoes mitosis, the number of mutations that occurs is Poisson distributed and an average of 8 mutations occur. What is the probability that no more than 1 mutation occurs when two cells divide?

Solution: This is a Poisson process with $\lambda = 8$ for one cell, but if two cells divide, we expect an average of $\lambda = 16$ mutations to occur. We want to know $P(X \leq 1) = P(X = 0) + P(X = 1) = f(0) + f(1) = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} = e^{-16} + 16e^{-16} = 17e^{-16}$.

9. The number of chocolate chips in a cookie is Poisson distributed with an average of 5 chocolate chips. What is the probability that you pick up a cookie with 10 chocolate chips in it?

Solution: This is a Poisson distribution with $\lambda = 5$. We want to calculate $f(10) = \frac{\lambda^{10} e^{-\lambda}}{10!} = \frac{5^{10} e^{-5}}{10!}$.

10. The number of errors on a page is Poisson distributed with approximately 0.2 errors per 50 pages of a book. What is the probability that a novel of 300 pages contains no errors?

Solution: This is a Poisson distribution and the probability that we have 300 pages with no errors means that the first 50 pages has no errors, the second 50 and so on to the last 50 all have no errors. So the answer is $f(0)^6 = e^{-6\lambda} = e^{-1.2}$.

Another way to do this is noting that if we average 0.2 error per 50 pages, then over a novel of 300 pages, we should expect $300/50 \cdot 0.2 = 1.2$ errors. Thus, the probability of having no errors with $\lambda = 1.2$ is $e^{-\lambda} = e^{-1.2}$.

11. Approximately 4 people are born every second. What is the probability that in a minute, there are 80 people born?

Solution: Since approximately 4 people are born every second and there are 60 seconds in a minute, approximately $4 \cdot 60 = 240$ people are born every minute. Thus, this is Poisson distribution with $\lambda = 240$ and we want to calculate $f(80) = \frac{\lambda^{80}e^{-\lambda}}{80!} = \frac{240^{80}e^{-240}}{80!}$.

2 Expected Value and Variance

2.1 Concepts

12.

Distribution	PMF	$E(X)$	Variance
Uniform	If $\#R(X) = n$, then $f(x) = \frac{1}{n}$ for all $x \in R(X)$.	$\sum_{i=1}^n \frac{x_i}{n}$	$\sum_{i=1}^n \frac{(x_i-\mu)^2}{n}$
Bernoulli Trial	$f(0) = 1-p$, $f(1) = p$	p	$Var(X) = p(1-p)$
Binomial	$f(k) = \binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Geometric	$f(k) = (1-p)^k p$	$\frac{1-p}{p}$	$Var(X) = \frac{1-p}{p^2}$
Hyper-Geometric	$f(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$
Poisson	$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

The **Expected Value** is the weighted average of all the values the random variables can take on. By definition, it satisfies some properties:

- $E[c] = c$
- $E[cX] = cE[X]$
- $E[X+Y] = E[X] + E[Y]$ for **all** random variables
- $E[XY] = E[X]E[Y]$ for **independent** random variables.

The **Variance** is defined as $Var(X) = E((X-\mu)^2)$. An easier form is $E(X^2) - E(X)^2$. It satisfies some properties:

- $Var(c) = 0$
- $Var(cX) = c^2 Var(X)$
- $Var(X+Y) = Var(X) + V(Y)$ for **independent** random variables.

2.2 Examples

13. I flip a fair coin 5 times. What is the expected number of heads I flip and what is the variance?

Solution: This is a binomial distribution with $p = \frac{1}{2}$ and $n = 5$. Then $E(X) = 5/2$ and $Var(X) = \frac{5}{4}$.

14. I roll two fair 6 sided die. What is the expected value of their product?

Solution: Let X be the first value I roll and Y be the second. The rolls are independent and so $E[XY] = E[X]E[Y] = 3.5 \cdot 3.5 = 12.25$.

15. I flip a fair coin 10 times. What is the expected number of pairs of consecutive heads I flip? (The sequence HHH has two pairs of consecutive heads)

Solution: Let X_{12} be 1 if the first and second flips are heads, $X_{23}, \dots, X_{910}$ defined similarly. Then the number of pairs of consecutive heads I flip is just $X_{12} + X_{23} + \cdots + X_{910}$. By linearity of expectation, we have that $E[X_{12} + X_{23} + \cdots + X_{910}] = E[X_{12}] + E[X_{23}] + \cdots + E[X_{910}]$. For each of these, the probability that $X_{12} = 1$ is $\frac{1}{4}$ and probability that it is 0 is $\frac{3}{4}$. So the expected value is $\frac{1}{4}$ and hence the expected number of consecutive heads is $\frac{9}{4}$.

2.3 Problems

16. True **FALSE** The expected value of a random variable X is the value such that the PMF at that point is the largest.

Solution: The PMF at that point may be 0! For example, there expected number of heads when flipping 5 coins is 2.5 but we cannot flip 2.5 heads.

17. True **FALSE** The expected value of a random variable X always exists.

Solution: This is false as seen in the homework.

18. While pulling out of a box of cookies, what is the expected number of cookies I have to pull out before I pull out an oatmeal raisin if 20% of cookies are oatmeal raisin? What is the variance?

Solution: This is a geometric distribution because I am counting the number of cookies I have to pull out before a success. The probability of success is $20\% = p = 1/5$. So the expected number of cookies I have to pull out is $\frac{1-p}{p} = 4$. The variance is $\frac{1-p}{p^2} = 4/(1/5) = 20$.

19. What is the expected number of aces I have when I draw 5 cards out of a deck?

Solution: Drawing cards out of a deck without replacement is the hypergeometric distribution. There are $N = 52$ cards total and $m = 4$ aces total. Then, we pull out $n = 5$ cards and so the expected number of aces is $\frac{mn}{N} = \frac{20}{52}$.

20. In a safari, safari-keepers have caught and tagged 300 rhinos. On a safari, out of the 15 different rhinos you see, there are 5 of them expected to be tagged. How many rhinos are there at the safari?

Solution: This is a hyper-geometric distribution because out of the N rhinos total and $m = 300$ tagged rhinos, you see that $n = 15$ rhinos that you see, there are 5 of them expected to be tagged. So $5 = E(X) = \frac{mn}{N} = \frac{300 \cdot 15}{N}$. So $N = \frac{300 \cdot 15}{5} = 900$.

21. In a class of 30 students, I split them up into 6 groups of 5 on Tuesday. Today, Thursday, I split them up again randomly. What is the expected number of people in your new group were in your old group on Tuesday?

Solution: We can think of this as a hypergeometric distribution where day 1, we "tag" or mark the 4 students that were in your group. Then on day 2, out of the 29 other students, you want to select 4 of them without replacement to be in your group. This is a hyper-geometric distribution with $N = 29, n = 4$. Then $m = 4$ because there are 4 students that were in your group before. So, the expected number of people in your new group who were in your old group is $\frac{mn}{N} = \frac{4 \cdot 4}{29} = \frac{16}{29}$.

22. In a class of 30 students, I split them up into 6 groups of 5. What is the expected number of days of splitting them up randomly into new groups of 5 before I split them up into the same groups again (assume that the groups are indistinguishable)?

Solution: First we count the number of ways to split up 30 people into 6 groups of 5. First we choose the first group and that is $\binom{30}{5}$ ways, then the second is $\binom{25}{5}, \ldots$, and the last group is $\binom{5}{5}$ ways. Finally, the 6 groups are indistinguishable which means it doesn't matter which way we order the groups so we need to divide by 6!. Thus, the total number of ways to make the groups is

$$N = \frac{\binom{30}{5}\binom{25}{5}\binom{20}{5}\binom{15}{5}\binom{10}{5}\binom{5}{5}}{6!}.$$

Since each of these group splitting is chosen randomly, the probability of splitting them up the same way is $p = \frac{1}{N}$. And the expected number of days I need before splitting them up in the same way is $\frac{1-p}{p} = N - 1$.

2.4 Extra Problems

23. While pulling out of a box of cookies, what is the expected number of cookies I have to pull out before I pull out an oatmeal raisin if 15% of cookies are oatmeal raisin? What is the variance?

Solution: This is a geometric distribution because I am counting the number of cookies I have to pull out before a success. The probability of success is $15\% = p = 3/20$. So the expected number of cookies I have to pull out is $\frac{1-p}{p} = \frac{17}{3}$. The variance is $\frac{1-p}{p^2} = \frac{340}{9}$.

24. What is the expected number of kings I have when I draw 8 cards out of a deck?

Solution: Drawing cards out of a deck without replacement is the hypergeometric distribution. There are $N = 52$ cards total and $m = 4$ kings total. Then, we pull out $n = 8$ cards and so the expected number of kings is $\frac{mn}{N} = \frac{32}{52}$.

25. In a safari, safari-keepers have caught and tagged 100 rhinos. On a safari, out of the 20 different rhinos you see, there are 8 of them expected to be tagged. How many rhinos are there at the safari?

Solution: This is a hyper-geometric distribution because out of the N rhinos total and $m = 100$ tagged rhinos, you see that $n = 20$ rhinos that you see, there are 8 of them expected to be tagged. So $8 = E(X) = \frac{mn}{N} = \frac{100 \cdot 20}{N}$. So $N = \frac{100 \cdot 20}{8} = 250$.