Chapter 4: Discrete Distributions

Binomial Distribution

Example 5: Suppose a random sample of 15 patients receive certain treatment. Each patient shows improvement in their health with probability 0.7. Find the probability that more than 10 people in the sample will show improvement.

Bernoulli experiment/trial $\Rightarrow$ experiment that has 2 possible outcomes "success" & "failure"

Binomial experiment: n independent Bernoulli trials, each has probability of "success" p

- fixed # of trials, n
- independent trials
- each trial results $\rightarrow$ pass, $\rightarrow$ fail
- prob. of success is const in each trial, p

$\Rightarrow$ properties of binomial experiment

Random variable X = the number of successes in the n trials,

X = # success in the n trials

$\boxed{\Rightarrow X \sim \text{Binom}(n,p)}$

X is Binomial random variable with parameters n & p

n is the number of trials

NOTES:

1. Bernoulli is a special case of Binomial distribution (when $n = 1$).

$X \sim \text{Binom}(1,p)$

or

$X \sim \text{Bernoulli}(p)$

$\Rightarrow$ Bernoulli implies $n=1$.

2. X = the number of successes in n independent Bernoulli trials. For ith trial define

$$X_i = \begin{cases} 1 & \text{if the ith trial is success} \\ 0 & \text{if the ith trial is failure} \end{cases}, \quad i = 1,2,\dots,n$$

Then $X = X_1 + X_2 + \cdots + X_n$

3. Consider Y = the number of "failures" in the n trials

Generally, X = # of success in n trials | $\boxed{X \sim \text{Binom}(n, p)}$

so, let Y = failures in n trials

$\boxed{\Rightarrow X + Y = n \; ; \; Y \sim \text{Binom}(n, (1-p))}$

Properties of Binomial distribution

Support: $S_X = \Rightarrow S_X = \{0, 1, 2, \dots, n\}$

Pmf: $p(k) =$

$P_X(k) = P(X = k) = \boxed{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}$

// $\therefore (n-k)$ failures

//exactly k successes in n trials

Cdf: no closed-form formula for $F(k)$ → i) Use binomial table

$\boxed{\text{No formula!}}$ → ii) softwares

$\therefore$ No cdf for this class

Example 5 ⇒ fixed # patients, 15 ⇒ $n=15$
⇒ independent patients / trials
⇒ each patient < success / failure ⇒ binomial ✓
⇒ Prob. to improve is the same for each patient, $p=0.7$

Let X = # of patients among the 15 who improve
$X \sim \text{Binom}(n=15, p=0.7)$

$\boxed{P(X>10)\,?}$ $X \sim \text{Binom}(15, 0.7)$

i) Support ⇒ $S_X = \{0,1,2,\ldots n\}$

ii) pmf ⇒ $p_X(k) = P(X=k) = \boxed{\binom{n}{k}\cdot p^k \cdot (1-p)^{n-k}}$

// ∴ $(n-k)$ failures
// exactly k successes in n trials

iii) cdf ⇒ $\boxed{\text{No formula !!}}$ → i) Use binomial table
→ ii) Softwares

$\therefore P(X>10) = p_X(11) + p_X(12) + \ldots p_X(15) = \binom{15}{11}(0.7)^{11}(1-0.7)^4 + \binom{15}{12}(0.7)^{12}(0.3)^3 + \ldots - \binom{15}{15}(0.7)^{15}(0.3)^0 = \boxed{0.515\ldots}$

⊛ Expected # patients to improve $\to X$

$E(X) = np = 15 \cdot (0.7) = 10.5$

⊛ Don't round up expectations! keep decimals same as in n or p or 1 more for more precision.

$Var(X) = 15 \cdot (0.7)(1-0.7) = 3.15 = n \cdot p \cdot (1-p)$

⊛ Expected # patients to improve → X

$E(X) = np = 15 \cdot (0.7) = 10.5$

⊛ Don't round up expectations! keep decimals same as in n or p or 1 more for more precision.

Mean and Variance of Binomial distribution

$Var(X) = 15 \cdot (0.7)(1-0.7) = 3.15 = n \cdot p \cdot (1-p)$

Example 6: A hip joint replacement part is stress-tested in a lab. The probability for a part to successfully complete the test is 0.80. A random sample of seven parts is tested.

a) What is the probability that exactly 2 of the seven parts successfully complete the test?

b) What is the probability that none of the tested parts fail the test?

- fixed # parts (trials), $n=7$
- independent parts
- each part → complete the test / → doesn't
- prob. to complete the test for each part is same, $p=0.8$

a) $P(X=2) = p_X(2) = \binom{7}{2} 0.8^2 \, 0.2^5 = \boxed{0.0043}$

b) $P(X=7) = p_X(7) = \binom{7}{7} (0.8)^7 (0.2)^0 = \boxed{0.2097}$ // ∵ all pass

c) Probability that at most 5 parts successfully complete the test

$P(X \leq 5) = 1 - p_X(6) - p_X(7) = \boxed{0.5767}$

Poisson Distribution

→ when some event is happening & is time dependent

Let X be the number of occurrences of an event in a given time period (area, space).

$\boxed{\text{\# of occurences of an event}}$

Poisson Process

We have a Poisson process if events occur in time/space according to the following rules:

- independent occurences
- occurences happen at constant rate λ, $\lambda > 0$
- probability of 2 or more occurences in a very small time period is negligible ≈ 0

Suppose events occur according to a Poisson process at a rate of λ events per unit time interval, $\lambda > 0$.

Let $X =$ the number of events that occur in a unit time interval.

X is a Poisson's random variable with parameters λ.

$$\boxed{X \sim Pois(\lambda)}$$

Properties of Poisson Distribution

Support: $S_X = \{0, 1, 2, 3 \ldots\}$

Pmf: $\boxed{p(k) = p(k) = \frac{\lambda^k e^{-\lambda}}{k!}}$ where $k = 0, 1, 2 \ldots$

$$\boxed{e = 2.718}$$

Cdf: no closed-form formula for $F(k)$

Mean: $E(X) =$ λ (λ = rate of occurences)

Variance: $Var(X) =$ λ (λ = also equal to avg no. of occurences)

Example 7: Suppose customers walk into a store according to a Poisson process with a mean of 4 customers per hour. Find the probability that between 6 and 9 customers will arrive during the next hour. $\Rightarrow t=1$ ∴ Base case formula

X is no. of occurences in unit interval of time

$$P(6 \le X \le 9) = p(6) + p(7) + p(8) + p(9)$$

and, $X \sim Pois(\lambda)$

$$\boxed{E(X) = 4 = \lambda} \quad \Rightarrow P(6 \le X \le 9) = \frac{4^6 e^{-4}}{6!} + \frac{4^7 e^{-4}}{7!} + \frac{4^8 e^{-4}}{8!} + \frac{4^9 e^{-4}}{9!}$$

$$\boxed{= 0.207}$$

What if we count the events that occur in a time interval of length **t**?

Suppose events occur according to a Poisson process with a mean rate of λ events **per unit time**. Let $X =$ the number of events that occur in an interval of length t (t time units).

$$\boxed{X \sim Pois(\lambda t)}$$

⊛ t is time interval for which the event takes place.

⊛ $$\boxed{\text{Pmf} \quad p_x = \frac{(\lambda t)^k e^{-\lambda t}}{k!}}$$ $k = 0, 1, 2, 3 \ldots$

⊛ $$\boxed{\begin{array}{l} E(X) = \lambda t \\ Var(X) = \lambda t \end{array}}$$

Example 7 (cont.): Suppose customers walk into a store according to a Poisson process with a mean of 4 customers per hour.

a) Find the probability that more than 8 customers will arrive during the next 150 minutes.
b) What is the expected number of customers that walk into the store in the next 150 min?

a) $\lambda = 4$; $Y =$ # of customers in next $\underbrace{150 \text{ min}}_{2.5\text{hr}}$ $(t = 2.5\text{hr})$

$$Y \sim Pois(\lambda t = 4 \cdot (2.5) = 10)$$

$$P(Y>8) = 1 - P(Y \le 8) = \sum_{k=9}^{\infty} \left(\frac{10^k e^{-10}}{k!} \right) = 1 - P_Y(0) - P_Y(1) \ldots P_Y(8)$$

$$\boxed{= 1 - \sum_{k=0}^{8} \frac{10^k e^{-10}}{k!}}$$

b) $E(Y) = 10 = \boxed{\lambda t}$

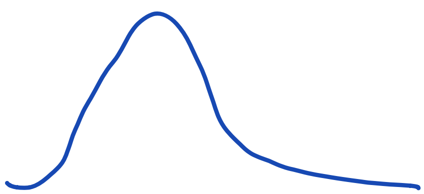
Poisson Approximation to Binomial Distribution

→ A binomial can be distributed as a Poisson if conditions are satisfied

Let $X \sim \text{Binom}(n, p)$.
We are interested in $P(a \le X \le b)$ (without loss of generality assume a and b are integers.)

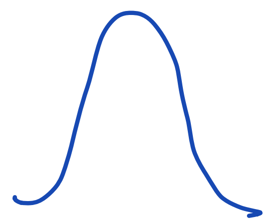
Shapes of Binomial pmf

p is small (close to 0)

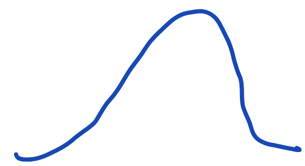


↳ always the skew

p is neither small nor large



p is large (close to 1)



Shape of Poisson pmf

The approximation is good if:

i) p is small ✓

ii) p is neither small nor large X

iii) p is large ✓ (indirectly for failures)

Check →

$n \geq 100$

p is small

$p \leq 0.01$, $np \leq 20$

Example 8: An article reports that 1 in 200 people carry the defective gene that causes inherited colon cancer. A random sample of 1000 individuals is selected. What is the approximate probability that at least 8 people in the sample carry the gene?

$X =$ # of people in the sample with the gene $\Rightarrow P(X \geq 8)$

i) fixed # of people (trials) $n = 1000$

ii) independent people

iii) each person has $\rightarrow$ the gene / $\rightarrow$ doesn't have the gene

iv) prob to carry, the gene in each person $\Rightarrow p = \frac{1}{200} = 0.005$

$\therefore \; X \sim \text{Binom}(n = 1000, p = 0.005)$

Check $\Rightarrow$ $n = 1000 \geq 100$ ✓

$p = 0.005 \leq 0.01$ ✓

$\boxed{\lambda = np} = 5 \leq 20$ ✓

$\Rightarrow$ Checking appropriateness for Poisson approx

$\Rightarrow \; X \sim \text{Pois}(5)$

$\Rightarrow P(X \geq 8) = 1 - P_{Pois(5)}(0) - P_{Pois(5)}(1) \cdots - P_{Pois(5)}(7)$

$\boxed{= 0.133}$

perform trials till rth success

Geometric and Negative Binomial Distributions

Example 9: Suppose Joe plays a dart game with his friends and doesn't quit until he hits the bull's-eye (a perfect hit). Assume that each time he throws the dart there is a 0.6 for him to get a perfect hit.

a) Find the probability that a least 5 throws are needed for him to get a perfect hit?
b) On average how many throws it takes for him to get a perfect hit?
c) What is the probability he will get 4 perfect hits in no more than 7 trials?

Negative Binomial experiment: perform <u>independent</u> Bernoulli trials, until the rth "success"

- independent trials
- each trial $\to$ success / failure
- prob of success is same for all trials, p
- perform trials until r^{th} success

X = the number trials until the rth success

X is negative binomial, r.v. .. with parameters r & p

$$\boxed{X \sim (\text{Neg Binom}(r, p))}$$

Properties of Negative Binomial:

Support: $S_X = \{r, r+1, r+2 \dots\}$ # trials

Pmf: $p(k) = P(X=k) = \binom{k-1}{r-1} \cdot p^r \cdot (1-p)^{k-r}, \quad k = r, r+1 \dots$

1st   2nd   . . . .   (k−1)th   kth

r−1 success | k−2 failures

Cdf: in general no closed-form formula for $F(k)$

Mean and variance:

$$\boxed{E(X) = \frac{r}{p}} \quad ; \quad \boxed{Var(X) = \frac{r(1-p)}{p^2}}$$

Special case: ⇒ Geometric Distribution

when $\boxed{r = 1}$ ⇒ X = no. of trials until first success,

$\boxed{X \sim NegBinom(r=1, p)}$

$\boxed{X \sim Geometric(p)}$

Support: $S_X =$

Pmf: $p(k) = P(X=k) = p(1-p)^{k-1}$

$k = 1, 2, 3$

Cdf: $F(k) = P(X \leq k) = 1-(1-p)^k$

Mean and variance: $\boxed{E(X) = \frac{1}{p}}$ $\boxed{Var(X) = \frac{1-p}{p^2}}$

Memoryless property:

If $X \sim Geom(p)$, then for any Z^+, m & n

$$\boxed{\begin{aligned} P(X > m+n \mid X > m) &= P(X > n) \\ P(X \leq m+n \mid X > m) &= P(X \leq n) \end{aligned}}$$

	# trials	# success
Binom	fixed	RV
Neg Binom ↓ until rth success	RV	fixed

Example 9 (cont.): Suppose Joe plays a dart game with his friends and doesn't quit until he hits the bull's-eye (a perfect hit). Assume that each time he throws the dart there is a 0.6 for him to get a perfect hit.

a) Find the probability that a least 5 throws are needed for him to get a perfect hit?
b) On average how many throws it takes for him to get a perfect hit?
c) What is the probability he will get 4 perfect hits in no more than 7 trials?
d) on avg. how many throws to get 4 " " ...

- perform throws until 1st perfect hit.
- independent throws
- each throws → perfect hit / o/w
- prob. to get a perfect hit is same for all throws, $P = 0.6$

X = # throws until 1st perfect hit

$X \sim \text{Neg Binom}(r=1, p=0.6)$ ✓

$X \sim \text{Geometric}(p = 0.6)$

a) $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - F_X(4) =$
$1 - (1 - (1-0.6)^4)$
$= \boxed{0.0256}$

or,

$1 - P_X(1) - P_X(2) - P_X(3) - P_X(4)$
$= 1 - 0.6(1-0.6)^{1-1} - 0.6(1-0.6)^{2-1} - 0.6(1-0.6)^{3-1} - 0.6(1-0.6)^{4-1}$

b) $E(X) = \frac{1}{0.6} = \boxed{1.67}$

$Var(X) = \frac{1-p}{p^2} = \frac{1-0.6}{(0.6)^2} = \boxed{1.11}$

c) Y = # throws until 4th perfect hit

$Y \sim \text{Neg.Bin}\,(r=4,\ p=0.6)$

$P(y \le 7) = P_y(4) + P_y(5) + P_y(6) + P_y(7)$

$= \binom{4-1}{4-1} 0.6^4 (1-0.6)^{4-4} + \binom{5-1}{4-1} (0.6)^4 (1-0.6)^{5-4}$

$+ \binom{6-1}{4-1} 0.6^4 (1-0.6)^{6-4} + \binom{7-1}{4-1} (0.6)^4 (1-0.6)^{7-4}$

$= \boxed{0.71}$

d) $E(y) = \frac{4}{0.6} = \boxed{6.67}$

Hypergeometric Distribution

Suppose of the total of N objects, there are m objects of type 1 and $N - m$ objects of type 2. We select n objects randomly without replacement.

N objects (total)

m of type I (= ques)

$N-m$ of type II (!ques)

$n \rightarrow$ sample chosen

X = # of type I objects in sample

X is a hypergeometric R.V

$X \sim \text{Hypergeom}(N, m, n)$

$$\boxed{\circledast \ \text{pmf} = P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}}$$

$$\boxed{\circledast \ \text{pmf} = \begin{cases} 0 \le k \le m \\ 0 \le n-k \le N-m \end{cases}}$$

$$\boxed{\text{support } 0 \le k \le m \Rightarrow \max(0, n-N+m) \le k \le \min(n, m)}$$

Example 10: A city has 25 buses, and 8 of these buses have cracks on the underside of the main frame. A sample of 5 buses is randomly selected for inspection.

a) What is the probability that of the 5 buses exactly 4 have the cracked frame?

b) What is the probability that of the 5 buses at least 4 have the cracked frame?

$X \sim \text{Hypergeometric}(N=25, m=8, n=5)$

a) $$P(X=4) = \frac{\binom{8}{4}\binom{17}{1}}{\binom{25}{5}} = \boxed{0.022}$$

b) $$P(X \geq 4) = P(X=4) + P(X=5)$$
$$= 0.022 + \frac{\binom{8}{5}\binom{17}{0}}{\binom{25}{5}}$$

Revision Notes

- Bernoulli $\Rightarrow$ 1 trial ; p $\Rightarrow$ $X \sim Binom(1, p)$

- Binomial $\Rightarrow$ n trials $\Rightarrow$ $X \sim Binom(n, p)$

- Probability $P(X > a) = \sum_{i=a+1}^{n} \binom{n}{i} (p)^{i} (1-p)^{n-i}$

  $i \Rightarrow$ // successful ones

- $$\boxed{\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x}$$

- Mention & check all the properties first before proceeding with calculation.

- $\boxed{E(X) = np}$ & $\boxed{Var(X) = np(1-p)}$ in case of Bernoulli

- Normal Poisson $X \sim Pois(\lambda)$ $\Rightarrow$ $\boxed{p(k) = \frac{\lambda^k e^{-\lambda}}{k!}}$; $\boxed{E(X) = Var(X) = \lambda}$

  $X \sim Pois(\lambda t)$ $\Rightarrow$ $\boxed{p(kt) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}}$; $\boxed{E(X) = Var(X) = \lambda t}$

  where $E(X)$ = rate of occurences & $Var(X)$ = avg. no. of occurences

Example 7 (Additional Practice):

Suppose customers walk into a store according to a Poisson process with a mean of 4 customers per hour.

(c) What should be the length of the time period so that the probability of no customers coming to the store during this time period is at most 0.35?

NOTE: For better learning, make a serious effort to solve this question on your own before looking at the solution on next page!

SOLUTION:

The customers arrive at the rate $\lambda = 4$ per hour. We choose 1 hour to be the time unit.

Consider X = the number of customers arriving in t hours (= in t time units). Then $X \sim \text{Pois}(4t)$.

We are given that probability of no customers coming to the store during this time period is at most 0.35, i.e. $P(X = 0) \leq 0.35$, or

$$P(X = 0) = p_X(0) = \frac{(4t)^0 e^{-4t}}{0!} = e^{-4t} \leq 0.35.$$

and

$$e^{-4t} \leq 0.35 \quad \Leftrightarrow \quad -4t \leq \ln(0.35) \quad \Leftrightarrow \quad t \geq -\frac{\ln(0.35)}{4} = 0.262,$$

i.e. at least 0.262 hour is needed for the probability of no customers coming to the store during this time period to be at most 0.35

NOTE: We could have chosen another time unit.

The key is to keep units consistent!

For practice purposes, suppose we choose the time unit to be, say, 1 minute. Then the rate per minute is $\frac{4}{60}$.

If now X = the number of customers arriving in t minutes (= in t time units), then $X \sim \text{Pois}\left(\frac{4}{60}t\right)$ and

$$P(X = 0) = \frac{\left(\frac{4}{60}t\right)^0 e^{-\frac{4}{60}t}}{0!} = e^{-\frac{4}{60}t} \leq 0.35$$

$$\Leftrightarrow \quad -\frac{4}{60}t \leq \ln(0.35) \quad \Leftrightarrow \quad t \geq -\frac{60}{4}\ln(0.35) = 15.7.$$

And note that 15.7 minutes is the same time period as 0.262 hours (just expressed in different units).

Example 6 (Additional Practice):

A hip joint replacement part is stress-tested in a lab. The probability for a part to successfully complete the test is 0.80. A random sample of seven parts is tested.

If there is a $100 fine for any part that doesn't pass the test, what is the expected fine for the sample?

NOTE: For better learning, make a serious effort to solve this problem on your own before looking at the solution on next page!

SOLUTION:

Recall that in class we introduced X = the number of parts in the sample of 7 that complete the test, and discussed that $X \sim \text{Binom}(n = 7, p = 0.8)$.

Since X represents the number of parts in the sample that pass the test, $7 - X$ is the number of parts that don't pass the test. Then the fine

$$F = 100(7 - X) = 700 - 100X,$$

which is a linear function of X and so the expected fine

$$E(F) = E\,(700 - 100X) = 700 - 100E(X).$$

Since $X \sim \text{Binom}(7, 0.8)$, we know $E(X) = 7 * 0.8 = 5.6$ and then the expected fine

$$E(F) = 700 - 100 * 5.6 = 140\ \$$$