STAT 320: Principles of Probability

Unit 5(PART:B)

A Few Discrete Random Variables

United Arab Emirates University

Department of Statistics

Outline

Binomial Distribution

- A quality controlling team in TV manufacturer evaluates a products before they are send to market. They are interested in idensifying the number of defective products and the resources that should be optimally allocated to mitigate the defectives items.
- A broker in a stock exchange speculates how many shares out of 10 newly introduced shares will go up next day.
- A airline company is interested in identifying the number of last minute cancellations that may take place.
- A car insurance company have sold 2000 insurance policies on a specific new cars. They need to estimate the funts that should be made available/reserved to compensate the losses that may occur to the insured cars during the next one year.

A Bernoulli Trial/ Experiment

- The random experiment has only two outcomes. Namely SUCCESS, and FAILURE
- Events corresponding to the successive trials/experiemnts are statistically independent.
- All such trials/experiements have the **same chance/probability of success.**

If a sequence of n independent Bernoulli trials is performed under the same condition, then the random variable that records the **total number of successes** is called the **Binomial Random variable.**

Binomial Distribution

Definition (Binomial Experiment)

An experiment is called a Binomial experiment if it satisfies the following 4 conditions:

- The experiment consists of n Bernoulli trials.
- Each trial results in a success (S) or a failure (F).
- The trials are independent.
- The probability of a success, π, is fixed throughout n trials.

Binomial Distribution Binomial(n, π)

1. Given a Binomial experiment consisting of n Bernoulli trials with success probability π, the Binomial random variable X associated with this experiment is defined as the number of successes among the n trials.
2. The random variable X has the Binomial Distribution with parameters n and π; denoted by $X \sim Binomial(n, \pi)$.
3. The behavior of Binomial Distribution with different n and π.

Binomial Distribution Binomial(n, π)

Definition (Binomial Distribution)

Let $\pi \in (0, 1)$, then the probability mass function of Binomial(n, π) is given by

$$\boxed{p(x) := \binom{n}{x} \pi^x (1-\pi)^{n-x}}, \text{ for } x \in \mathbb{S}_x, \text{ where } \mathbb{S}_x = \{0, 1, \ldots, n\}$$

Let $X \sim \text{Binomial}(n, \pi)$

Mean

$$E(X) = n\pi$$

Variance

$$\text{VAR}(X) = n\pi(1-\pi)$$

MGF

$$\text{M}_X(t) = \left(1 - \pi + pe^t\right)^n$$

Distribution	Support $\mathbb{S}_X$	pmf $p_X(x)$	Mean $E(X)$	Variance $\text{Var}(X)$	mgf $M_X(t)$
$\text{Binomial}(n, \pi)$	$\{0, 1, \ldots, n\}$	$\binom{n}{x}\pi^x(1-\pi)^{n-x}$	$n\pi$	$n\pi(1-\pi)$	$\left(1 - \pi + pe^t\right)^n$

Reminder from Unit1: Binomial Series

Let $x \in \mathbb{R}$ be any real number, and $n \in \mathbb{Z}_+$ be any positive integer, then

$$(1+x)^n = \boxed{\sum_{i=0}^{n} \binom{n}{i} x^i}$$

$$(1+x)^n = \boxed{1 + \binom{n}{1} x + \binom{n}{2} x^2 + \cdots + \binom{n}{n-1} x^{n-1} + \binom{n}{n} x^n}$$

Let $x \in \mathbb{R}$ be any real number, and $n \in \mathbb{Z}_+$ be any positive integer, then

$$(a+b)^n = \boxed{\sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}}$$

$$(a+b)^n = \boxed{b^n + \binom{n}{1} ab^{n-1} + \binom{n}{2} a^2 b^{n-2} + \cdots + \binom{n}{n-1} a^{n-1} b + \binom{n}{n} a^n}$$

Expected Value of Binomial Distribution

$$
\begin{aligned}
E(X) &:= \sum_{y \in \mathbb{S}_X} y \, p_x(y) \\
&= \sum_{y=0}^{n} y \binom{n}{y} \pi^y (1-\pi)^{n-y} \\
&= (1-\pi)^n \sum_{y=0}^{n} y \binom{n}{y} \left(\frac{\pi}{1-\pi} \right)^y \\
&= (1-\pi)^n \frac{np}{(1-\pi)^n} \\
&= np
\end{aligned} \tag{1}
$$

Expected Value of Binomial Distribution

$$\begin{aligned}
E(X^2) &:= \sum_{y \in \mathbb{S}_X} y^2 \, p_X(y) \\
&= \sum_{y=0}^{n} y^2 \binom{n}{y} \pi^y (1-\pi)^{n-y} \\
&= (1-\pi)^n \sum_{y=0}^{n} y^2 \binom{n}{y} \left(\frac{\pi}{1-\pi}\right)^y \\
&= (1-\pi)^n \frac{mp + n(n-1)\pi^2}{(1-\pi)^n} \\
&= np + n(n-1)\pi^2 \qquad (2)
\end{aligned}$$

$Var(X) = E(x^2) - (E(X))^2 = np + n(n-1)\pi^2 - n^2p^2 = np - np^2 = np(1-\pi)$.

Expected Value of Binomial Distribution

$$
\begin{aligned}
M_X(t) &:= \sum_{y \in \mathbb{S}_X} e^{ty}\, p_x(y) \\
&= (1-\pi)^n \sum_{y=0}^{n} e^{ty} \binom{n}{y} \left(\frac{\pi}{1-\pi}\right)^y \\
&= (1-\pi)^n \sum_{y=0}^{n} \binom{n}{y} \left(\frac{pe^t}{1-\pi}\right)^y \\
&= (1-\pi)^n \sum_{y=0}^{n} e^{ty} \binom{n}{y} \left(\frac{\pi}{1-\pi}\right)^y \\
&= (1-\pi)^n \left(1 + \frac{pe^t}{1-\pi}\right)^n = \left(1 - \pi + pe^t\right)^n
\end{aligned} \qquad (3)
$$

Suppose: $n = 100, \pi = 0.2$

$$p(x) = \binom{n}{x} \pi^x (1-\pi)^x$$

For $x \in \{0, 1, \ldots, 100\}$

$$E(X) = n\pi = 100 \times 0.2 = 20.$$

$\pi = 0.2$

x	p(x)
0	2.03704E-10
1	5.09259E-09
2	6.30208E-08
3	5.1467E-07
4	3.12019E-06
5	1.49769E-05
6	5.92835E-05
7	0.000199023
8	0.000578411
9	0.001478163
10	0.00336282
11	0.006878495
12	0.012753877
13	0.021583484
14	0.033531484
15	0.048061794
16	0.06383207
17	0.07885138
18	0.090898119
19	0.098074286
20	0.099300215

x	p(x)
21	0.094571633
22	0.084899534
23	0.07198004
24	0.05773399
25	0.043877833
26	0.031642668
27	0.021681087
28	0.014131423
29	0.008771228
30	0.005189643
31	0.002929637
32	0.001579258
33	0.000813557
34	0.000400796
35	0.000188947
36	8.52885E-05
37	3.68815E-05
38	1.52864E-05
39	6.07537E-06
40	2.31624E-06

x	p(x)
41	8.474E-07
42	2.976E-07
43	1.0035E-07
44	3.2501E-08
45	1.0111E-08
46	3.0224E-09
47	8.6814E-10
48	2.3964E-10
49	6.3579E-11
50	1.6213E-11
51	3.9737E-12
52	9.3611E-13
53	2.1195E-13
54	4.6118E-14
55	9.643E-15
56	1.9372E-15
57	3.7385E-16
58	6.929E-17
59	1.2331E-17
60	2.1066E-18

x	p(x)
61	3.4534E-19
62	5.4308E-20
63	8.1893E-21
64	1.1836E-21
65	1.6389E-22
66	2.1727E-23
67	2.7564E-24
68	3.3442E-25
69	3.8773E-26
70	4.2928E-27
71	4.5346E-28
72	4.5661E-29
73	4.3785E-30
74	3.9939E-31
75	3.4614E-32
76	2.8465E-33
77	2.2181E-34
78	1.6351E-35
79	1.1384E-36
80	7.4705E-38

x	p(x)
81	4.611E-39
82	2.671E-40
83	1.448E-41
84	7.328E-43
85	3.448E-44
86	1.504E-45
87	6.049E-47
88	2.234E-48
89	7.53E-50
90	2.301E-51
91	6.321E-53
92	1.546E-54
93	3.325E-56
94	6.189E-58
95	9.773E-60
96	1.273E-61
97	1.312E-63
98	1.004E-65
99	5.071E-68
100	1.268E-70

$$SD(X) = \sqrt{n\pi(1-\pi)} = \sqrt{100 \times 0.2(1-0.2)} = \sqrt{16} = 4.0$$

Example

Example : Five fair coins are flipped. If the outcomes are assumed independent.

1. Find the probability mass function of the number of heads obtained.
2. Find the probability that at least 3 heads are obtained.
3. Find the probability that at most 2 heads are obtained.

Example

Example : Five fair coins are flipped. If the outcomes are assumed independent.

1. Find the probability mass function of the number of heads obtained.
2. Find the probability that at least 3 heads are obtained.
3. Find the probability that at most 2 heads are obtained.

Solution: Let X = The number of heads in 5 tossed coins. $X \sim Binomial(n = 5, \pi = 0.5)$.

1. $P(X = 0) = 0.5^5 = 0.0313$
2. $P(X = 1) = \binom{5}{1}0.5^5 = 0.1563$
3. $P(X = 2) = \binom{5}{2}0.5^5 = 0.3125$
4. $P(X = 3) = \binom{5}{3}0.5^5 = 0.3125$
5. $P(X = 4) = \binom{5}{4}0.5^5 = 0.1563$
6. $P(X = 0) = \binom{5}{5}0.5^5 = 0.0313$

Example

Example : It is known that screws produced by a certain company will be defective with probability .01, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace? Use the Binomial Calculator or Statistical Tables.

Example

Example : The following gambling game, known as the wheel of fortune (or chuck-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, i = 1; 2; 3, then the player wins i units; if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player?

Outline

Poisson Distribution

- Number of calls received by a customer desk in an hour.
- Number of imperfections in every square-meter of a glass panel used for making LCD TV.
- Number of robot malfunctions per day in an assembly line.
- Number of car accidents occurs during a year.

Poisson Distribution

The Poisson distribution models the number of occurrences of an event when there is a known average rate per unit time or space λ.

Definition (Poisson Distribution)

The requirements for a Poisson distribution are that:

1. no two events can occur simultaneously,
2. events occur independently in different intervals, and
3. the expected number of events in each time interval remain constant.

Poisson Distribution: pmf, Expected Value

The Poisson distribution models the number of occurrences of an event when there is a known average rate per unit time or space λ.

Definition (Poisson Distribution: pmf, Expected Value)

The requirements for a Poisson distribution are that:

1. The probability mass function of Poisson(λ) is given by

$$\boxed{p(x) = \frac{e^{-\lambda}\lambda^x}{x!}} \quad \text{for } x = 0, 1, 2, 3, \ldots$$

2. If $X \sim Poisson(\lambda)$, then $E(X) = \lambda$, and $\mathrm{Var}(X) = \lambda$.

Let $X \sim \text{Poisson}(\lambda)$

Mean

$E(X) = \lambda$

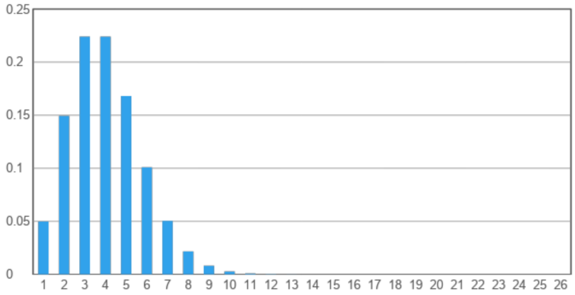
Variance

$\text{VAR}(X) = \lambda$

MGF

$\text{M}_x(t) = e^{\lambda e^t - \lambda}$

Distribution	Support $\mathbb{S}_x$	pmf $p_x(x)$	Mean $E(X)$	Variance $\text{Var}(X)$	mgf $M_x(t)$
Poisson(λ)	$\{0, 1, 2, \ldots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$e^{\lambda e^t - \lambda}$


0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

Reminder from Unit1: Exponential Series e^x or $(\exp(x))$

Definition (Exponential Series)

For any real number $x \in \mathbb{R}$, the exponential series e^x (or sometimes denoted as $\exp(x)$) is defined as,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

Expected Value of Binomial Distribution

$$
\begin{aligned}
M_X(t) &:= \sum_{y \in \mathbb{S}_X} e^{ty}\, p_x(y) \\
&= \sum_{y=0}^{\infty} e^{ty} \frac{e^{-\lambda}\lambda^{y}}{y!} \\
&= e^{-\lambda} \sum_{y=0}^{\infty} \frac{\left(\lambda e^{t}\right)^{y}}{y!} \\
&= e^{\lambda e^{t} - \lambda}
\end{aligned} \tag{4}
$$

- The number of customers arriving at a service counter within one-hour period.
- The number of typographical errors in a book counted per page.
- The number of email messages received at the technical support center daily.
- The number of traffic accidents that occur on a specific road during a month.

Poisson Process: Most Simple Version

1. The Number of Events Between the interval (can be time-interval or space-interval) $(s, t]$ follows

$$\text{Poisson}(\lambda \times (t - s))$$

   where $\lambda > 0$ denotes of rate of events per unit length of the interval.
2. Events pertaiing to the two distinct intervals are Statistically Independent
3. Rate of occurance of the events remain same for each of the subintervals with same length.

A Few Examples of Poisson Distribution

Example : Messages arrive at an electronic message center at random times, with an average of 9 messages per hour.

1. What is the probability of receiving exactly five messages during the next hour?
2. What is the probability that more than 10 messages will be received within the next two hours?

1. The number of messages received in an hour, X is modeled by Poisson distribution with $\lambda = 9$, i.e. $X \sim \text{Poisson}(9)$. $P(X = 5) = \frac{9^5 \exp(-9)}{5!}$
2. The number of messages received within a 2-hour period, Y is another Poisson distribution with $Y = (2)(9) = 18$, i.e. $Y \sim \text{Poisson}(18)$. $P(Y > 10) = 1 - P(Y \leq 10) = \ldots = 0.9696$

Outline

Geometric Distribution

Geometric Distribution

1. Suppose that independent trials, each having a probability π, $0 < \pi < 1$, of being a success, are performed until a success occurs.
2. Example: The first head in tossing coin several times.
3. Then, Geometric distribution models the number of trials performed until a success occurs.

Definition (Geometric Distribution)

The probability mass function of *Geometric*(π) is given by

$$p(x) = (1-\pi)^{x-1}\pi \text{ for } x = 1, 2, 3, \ldots,$$

Let $X \sim \text{Geometric}(\pi)$

Mean

$$E(X) = \frac{1}{\pi}$$

Variance

$$\text{VAR}(X) = \frac{1-\pi}{\pi}$$

MGF

$$M_X(t) = \frac{\pi e^t}{1-(1-\pi)e^t}$$

Distribution	Support $\mathbb{S}_X$	pmf $p_X(x)$	Mean $E(X)$	Variance Var(X)	mgf $M_X(t)$
Geometric(π)	$\{1, 2, \ldots\}$	$(1-\pi)^{x-1}\pi$	$\frac{1}{\pi}$	$\frac{1-\pi}{\pi}$	$\frac{\pi e^t}{1-(1-\pi)e^t}$

Reminder from Unit1: Geometric Series

Let $p \in \mathbb{R}$ be such that $|p| < 1$, then

$$\sum_{i=0}^{\infty} p^i = 1 + p + p^2 + p^3 + \cdots = \frac{1}{1-p}.$$

1. What is the value of $1 + 0.7 + (0.7)^2 + (0.7)^3 + \cdots =$
2. What is the value of $1 - 0.7 + (0.7)^2 - (0.7)^3 + \cdots =$

Expected Value of Geometric Distribution

$$\begin{aligned}
M_X(t) &:= \sum_{y \in \mathbb{S}_X} e^{ty}\, p_x(y) \\
&= \sum_{y=1}^{\infty} e^{ty}(1-\pi)^{y-1}\pi \\
&= \pi \sum_{z=0}^{\infty} e^{tz+t}(1-\pi)^z \\
&= pe^t \sum_{z=0}^{\infty} \left((1-\pi)e^t\right)^z \\
&= \frac{pe^t}{1-(1-\pi)e^t}
\end{aligned} \tag{5}$$

Geometric Distribution: Example

Example : Suppose that the probability of engine malfunction during any one-hour period is $\pi = 0.02$. Find the probability that a given engine will survive two hours.

Geometric Distribution: Example

Example : Suppose that the probability of engine malfunction during any one-hour period is $\pi = 0.02$. Find the probability that a given engine will survive two hours.

Solution:
Letting Y denote the number of one-hour intervals until the first malfunction, we have

$$\begin{aligned}
& P(\text{Survival for Next Two Hours}) \\
= \; & P(Y \geq 3) \\
= \; & 1 - P(Y \leq 2) \\
= \; & 1 - \sum_{y=1}^{2} p(y) \\
= \; & 1 - \{p(1) + p(2)\} \\
= \; & 1 - 0.02 - 0.98 \times 0.02 \\
= \; & 0.9604
\end{aligned}$$

Exercise Find the mean and standard deviation of Y.

Outline

Negative Binomial Distribution

1. Suppose that independent trials, each having probability π, $0 < \pi < 1$, of being a success are performed until a total of r successes is accumulated.
2. Example: The third head in tossing coin several times.
3. Then, Negative Binomial distribution models the number of trials performed until a the rth success occurs.

Definition (Negative Binomial Distribution)

The probability mass function of Negative Binomial RV, denoted by Negative-Binomial(r, π) is given by

$$p(x) = \binom{x-1}{r-1} \pi^r (1-\pi)^{x-r} \text{ for } x = r+1, r+2, r+3, \ldots,$$

Let $X \sim$ Negative-Binomial(r, π)

Mean

$$E(X) = \frac{r}{\pi}$$

Variance

$$\text{VAR}(X) = \frac{r(1-\pi)}{\pi}$$

MGF

$$\text{M}_X(t) = \left(\frac{\pi e^t}{1-(1-\pi)e^t}\right)^r$$

Distribution	Support $\mathbb{S}_X$	pmf $p_X(x)$	Mean $E(X)$	Variance $\text{Var}(X)$	mgf $M_X(t)$
Negative-Binomial(r, π)	$\{r+1, r+2, \ldots\}$	$\binom{x-1}{r-1}(1-\pi)^{x-r}\pi^r$	$\frac{r}{\pi}$	$\frac{r(1-\pi)}{\pi}$	$\left(\frac{\pi e^t}{1-(1-\pi)e^t}\right)^r$

Geometric Distribution: Example

Example : A machine produces 1% defective parts. Using the statistical calculator, calculate the probability that

1. 10 parts have to be selected until to get 2 defective parts.
2. Between 20 to 25 parts have to be selected to get 2 defective parts.

Geometric Distribution: Example

Example : A machine produces 1% defective parts. Using the statistical calculator, calculate the probability that

1. 10 parts have to be selected until to get 2 defective parts.
2. Between 20 to 25 parts have to be selected to get 2 defective parts.

Solution:

Exercise Find the mean and standard deviation of Y.

Outline

Miscellaneous Problems

Example : Approximately 10% of the glass bottles coming off a production line have serious flaws in the glass. If two bottles are randomly selected, find the mean and variance of the number of bottles that have serious flaws.

Example : Cars arrive at a toll both according to a Poisson process with mean 80 cars per hour. If the attendant makes a one-minute phone call, what is the probability that at least 1 car arrives during the call?

Example : Suppose that a lot of electrical fuses contains 5% defectives. If a sample of 5 fuses is tested, find the probability of observing at least one defective.

Example : An oil exploration firm is formed with enough capital to finance ten explorations. The probability of a particular exploration being successful is 0.10 . Assume the explorations are independent.

1. Find the mean and variance of the number of successful explorations.
2. Suppose the firm has a fixed cost of \$20,000 in preparing equipment prior to doing its first exploration. If each successful exploration costs \$30,000 and each unsuccessful exploration costs \$15,000, find the expected total cost to the firm for its ten explorations.

Example : A food manufacturer uses an extruder (a machine that produces bite-size cookies and snack food) that yields revenue for the firm at a rate of \$200 per hour when in operation. However, the extruder breaks down an average of two times every day it operates. If Y denotes the number of breakdowns per day, the daily revenue generated by the machine is $R = 1600 - 50Y^2$. Find the expected daily revenue for the extruder.

Example : A particular sale involves four items randomly selected from a large lot that is known to contain 10% defectives. Let Y denote the number of defectives among the 20 sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by $\text{Cost} = 3Y^2 + Y + 2$. Find the expected repair cost.

Example : In a certain population, it is known that 80% of the individuals have the Rhesus (Rh) factor present in their blood.

1. If 5 volunteers are randomly selected from the population, what is the probability that at least one does not have the Rh factor?
2. If five volunteers are randomly selected, what is the probability that at most four have the Rh factor?

Example : Consider rolling a fair dice multiple times untill the first 6 appears.

1. Find the expected number of throws required to get the first 6.
2. What is the probability that more then 8 throws are required to obtain the first 6?

Example : A tollbooth operator has observed that cars arrive randomly at a rate of 360 cars per hour

1. what is the probability that exactly two cars will come during a specific one-minute period?
2. Find the probability that 40 cars arrive between 10 am to 10:10 am
3. Find the expected number of cars between 10 am to 10:10 am

Questions?