Stat 154: Elementary Statistics

Jongyun Jung

Minnesota State University, Mankato

Ch 4: **Discrete Probability Distributions**

jongyun.jung@mnsu.edu

February 19, 2019

Overview

Definitions

- **Random Variable**: A Random Variable is a quantitative variable whose values are determined by chance.
- **Discrete Probability Distribution**: A discrete probability distribution satisfies three conditions:
  1. The random variable X is discrete, which can have only distinct or discrete values.
  2. The probability mass function (pmf) $0 \leq f(x) = P(X = x) = P(x) \leq 1$ for all x.
  3. $\sum_{\text{for all} x} f(x) = 1$, the sum of all individual probabilities is <u>one</u>.

Discrete Probability Distributions

- **Bernoulli Distribution**: If there are only two possible outcomes when a procedure is performed, it is called a Bernoulli experiment or a Bernoulli trial.
- We choose one of the two outcomes as a 'success' and the other one as a 'failure'. X can have only two possible values: 0 for 'failure' and 1 for 'success' and their probabilities are fixed but usually unknown.
- Then the distribution of X, with the probability mass function is

$$P(X = x) = f(x) = \begin{cases} 1 - p, \text{when} x = 0 \\ p, \text{when} x = 1 \end{cases} \tag{1}$$

- where $0 < p < 1$ is the probability of success and $1 - p$ is the probability of failure. X is called a Bernoulli random variable.

Discrete Probability Distributions

- **Binomial Distribution**: If a Bernoulli trial is repeated independently n times, we have a binomial experiment. X is the number of successes in these n independent Bernoulli trials.
- Then the distribution of X, with the probability mass function is

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, ..., n \qquad (2)$$

- is known as the binomial probability distribution. X is called a binomial random variable.

Discrete Probability Distributions

- **Poisson Distribution**: X is the number of occurrences in a unit interval or space when the rate of occurrence is a fixed value $\lambda \geq 0$. Then the distribution of X, with the probability mass function is

$$P(X = x) = f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0, 1, 2, ..., \tag{3}$$

- X is called a Poisson random variable.

Discrete Probability Distributions

In the manufacture of no-deposit-no-return bottles, 5 percent of the bottles are defective. What is the probability in a sample of seven bottles that there are

a. No defectives. $P(X = 0) = {}_7C_0(0.05)^0(1 - 0.05)^{7-0} = 0.6983.$

b. 3 defectives. $P(X = 3) = {}_7C_3(0.05)^3(1 - 0.05)^{7-3} = 0.0036.$

c. More than 3 defectives.

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - P(X = 0) - P(X - 1) - P(X = 2) - P(X = 3) \\ &= 1 - {}_7C_0(0.05)^0(1 - 0.05)^{7-0} - {}_7C_1(0.05)^1(1 - 0.05)^{7-1} \\ &\quad - {}_7C_2(0.05)^2(1 - 0.05)^{7-2} - {}_7C_3(0.05)^3(1 - 0.05)^{7-3} \\ &= 1 - 0.6983 - 0.2573 - 0.0406 - 0.0036 = 0.0002. \end{aligned}$$

d. At least one defective. $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.6983 = 0.3017.$

Figure: Bernoulli & Binomial Distribution

Discrete Probability Distributions

The transportation system at a new large airport is designed so that it will have one failure every 10 days. What is the probability that it will not fail on the "Grand Opening Day"?

Solution: Since it has one failure in every 10 days, it has average of 0.1 failures per day, which means $\lambda = 0.1$. Then

$$P(X = 0) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-0.1}(0.1)^{0}}{0!} = 0.9048.$$

Figure: Poisson Distribution

Other Situations

A box contains 100 marbles. Five of the marbles are purple and the rest are green. If eight marbles are drawn without replacement, what is the probability of obtaining exactly two purple marbles?

Solution: Since the drawings are without replacement, the composition of the box changes after each draw, so it is not a binomial experiment. The probability can be computed as,

$$\frac{{}_5C_2 \times {}_{95}C_6}{{}_{100}C_8} = \frac{\dfrac{5\cdot 4}{2\cdot 1} \times \dfrac{95\cdot 94\cdot 93\cdot 92\cdot 91\cdot 90}{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}}{\dfrac{100\cdot 99\cdot 98\cdot 97\cdot 96\cdot 95\cdot 94\cdot 93}{8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}} = 0.0467,$$

this computation also could be done using the combination formula given in Chapter 3.

Alternately, $P(2 \text{ in } 8 \text{ are purples}) = {}_8C_2 \times P(\text{PPGGGGGG})$, where P stands for purple and G stands for green. One choice is PPGGGGGG and this choice can happen in ${}_8C_2$ different ways. And

$$\begin{aligned} {}_8C_2 \times \text{P}(PPGGGGGG) &= 28 \times \frac{5}{100} \times \frac{4}{99} \times \frac{95}{98} \times \frac{94}{97} \times \frac{93}{96} \times \frac{92}{95} \times \frac{91}{94} \times \frac{90}{93} \\ &= 0.0467. \end{aligned}$$

Figure: Other Situations

Other Situations

A machine produces output with a fraction defective of 0.06. If a lot consists of 100 items, what is the probability that a sample of five items will contain one or fewer that are defective?

Solution: Since the sample size is finite, the sampling is without replacement as the composition changes after each draw. Let X be the number of defectives in a sample of 5 items. Among 100 items, there are 6 defective items. Then

$$P(X \le 1) = P(X=0) + P(X=1) = \frac{{}_6C_0 \times {}_{94}C_5}{{}_{100}C_5} + \frac{{}_6C_1 \times {}_{94}C_4}{{}_{100}C_5}$$

$$= \frac{1 \times \dfrac{94 \cdot 93 \cdot 92 \cdot 91 \cdot 90}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} + \dfrac{6}{1} \times \dfrac{94 \cdot 93 \cdot 92 \cdot 91}{4 \cdot 3 \cdot 2 \cdot 1}}{\dfrac{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = 0.9721.$$

Figure: Other Situations

Mean or Expected Value of a Discrete Random Variable

- The **Population mean** or the **expected value** for a discrete random variable can be computed as

$$\mu = E(X) = \sum_{\text{for all} x} x \cdot f(x) \tag{4}$$

- The **mean of the Bernoulli distribution** is p.
- The **mean for the binomial distribution** is np.
- The **mean for the Poisson distribution** is λ.

Variance and Standard Deviation of a Discrete Random Variable

- The **Population variance** for a discrete random variable is

$$\sigma^2 = E(x-\mu)^2 = \sum_{\text{for all} x} (x-\mu)^2 \cdot f(x) \tag{5}$$

- And $\sigma = \sqrt{\sigma^2}$ is the population standard deviation.
- The **variance of the Bernoulli distribution** is $p(1-p)$ and the standard deviation is $\sqrt{p(1-p)}$.
- The **variance of the Binomial distribution** is $np(1-p)$ and the standard deviation is $\sqrt{np(1-p)}$.
- The **variance of the Poisson distribution** is λ and the standard deviation is $\sqrt{\lambda}$.

Variance and Standard Deviation of a Discrete Random Variable

EXAMPLE 4.10 There are six people in an office. Each shows up independently for a picture with probability 0.9. Let X be the number of people that show up. What is the expected number of people that show up? What is the standard deviation of the number of people that show up?

Solution: The expected number of people that show up is $np = 6 \cdot 0.9 = 5.4$ and the standard deviation of the number of people that show up is

$$\sqrt{np(1-p)} = \sqrt{6 \cdot 0.9 \cdot (1-0.9)} = 0.73.$$

Figure: Other Situations

Variance and Standard Deviation of a Discrete Random Variable

EXAMPLE 4.11 On an average day in 2010, there were 203 traffic crashes on Minnesota roadways (from the 2010 Minnesota Motor Vehicle Crash Facts). What is the expected number of traffic crashes on any given day of 2010? What is the standard deviation of the number of the traffic crashes that occurred on Minnesota roadways on any given day of 2010?

Solution: Let X be the number of traffic crashes that occurred on Minnesota roadways in 2010. It is obvious that X follows a Poisson distribution with $\lambda = 203$. So the expected number of traffic crashes on any given day of 2010 is 203 and the standard deviation of the number of traffic crashes that occurred on any given day of 2010 is

$$\sqrt{\lambda} = \sqrt{203} \approx 14.2.$$

Figure: Other Situations

References

Mezbahur Rahman, Deepak Sanjel, Han Wu. Statistics Introduction, Revised Printing
KendallHunt