Sta 111 - Summer II 2017
Probability and Statistical Inference

4. Random variable and discrete distributions

Lu Wang

Duke University, Department of Statistical Science

July 5, 2017

Outline

1. Random Variables
    1. Expectation
    2. Variance
    3. Linear combinations of random variables
2. Binomial distribution
    1. Bernoulli random variable
3. Geometric distribution

- A *random variable* is a numeric quantity whose value depends on the outcome of a random event
  - We use a capital letter, like X, to denote a random variable
  - The values of a random variable are denoted with a lowercase letter, in this case x
  - For example, $P(X = x)$
- There are two types of random variables:
  - *Discrete random variables* often take only integer values
    - Example: a person's age in years, number of credit hours
  - *Continuous random variables* take real (decimal) values
    - Example: height of a person, cost of a book

- We are often interested in the average outcome of a random variable.
- We call this *expectation* (mean, expected value). For a discrete random variable, it is a weighted average of the possible outcomes

$$\mu = E(X) = \sum_{i=1}^{k} x_i \, P(X = x_i)$$

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. Write the probability model for your winnings, and calculate your expected winning.

Event	X	$P(X)$	$X\,P(X)$
Heart (not ace)	1	$\frac{12}{52}$	$\frac{12}{52}$
Ace	5	$\frac{4}{52}$	$\frac{20}{52}$
King of spades	10	$\frac{1}{52}$	$\frac{10}{52}$
All else	0	$\frac{35}{52}$	0
Total			$E(X) = \frac{42}{52} \approx 0.81$

- We are also often interested in the variability in the values of a random variable.
- We call this *variance*. For a discrete random variable, the variance is defined as

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= \sum_{i=1}^{k} (x_i-\mu)^2 P(X=x_i) \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

- *Standard deviation* is defined as

$$\sigma = SD(X) = \sqrt{Var(X)}$$

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\,P(X)$	$(X - E(X))^2$	$P(X)\;(X - E(X))^2$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
		$E(X) = 0.81$		$V(X) = 3.4246$ $SD(X) = \sqrt{3.4246} = 1.85$

Linear combinations

- A *linear combination* of random variables X and Y is given by

$$aX + bY$$

  where a and b are some fixed numbers.
- The *expectation* of a linear combination of random variables is given by

$$E(aX + bY) = aE(X) + bE(Y)$$

- The *variance* of a linear combination of two **independent** random variables is calculated as

$$V(aX + bY) = a^2 V(X) + b^2 V(Y)$$

- The standard deviation of the linear combination is the square root of the variance.

Note: If the random variables are not independent, the variance calculation gets a little more complicated and is beyond the scope of this course.

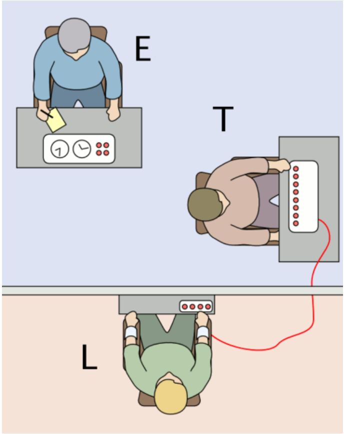
Adding or multiplying?

A company has 5 Lincoln Town Cars in its fleet. Historical data show that annual maintenance cost for each car is on average \$2,154 with a standard deviation of \$132. What is the mean and the standard deviation of the total annual maintenance cost for this fleet?

Note that we have 5 cars each with the given annual maintenance cost $(X_1 + X_2 + X_3 + X_4 + X_5)$*, not one car that had 5 times the given annual maintenance cost* $(5X)$*.*

$$
\begin{aligned}
E(X_1 + X_2 + X_3 + X_4 + X_5) &= E(X_1) + E(X_2) + \cdots + E(X_5) \\
&= 5 \times E(X) = 5 \times 2,154 = \$10,770 \\
Var(X_1 + X_2 + X_3 + X_4 + X_5) &= Var(X_1) + Var(X_2) + \cdots + Var(X_5) \\
&= 5 \times V(X) = 5 \times 132^2 = \$87,120 \\
SD(X_1 + X_2 + X_3 + X_4 + X_5) &= \sqrt{87,120} = 295.16
\end{aligned}
$$

- Stanley Milgram, a Yale University psychologist, conducted a series of experiments on obedience to authority starting in 1963.
- Experimenter (E) orders the teacher (T), the subject of the experiment, to give severe electric shocks to a learner (L) each time the learner answers a question incorrectly.
- The learner is actually an actor, and the electric shocks are not real, but a prerecorded sound is played each time the teacher administers an electric shock.



http://en.wikipedia.org/wiki/File:

Milgram_Experiment_v2.png

- These experiments measured the willingness of study participants to obey an authority figure who instructed them to perform acts that conflicted with their personal conscience.

- Milgram found that about 65% of people would obey authority and give such shocks.

- *Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?*

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underbrace{0.35}_{\text{(A) refuse}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 2: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.35}_{\text{(B) refuse}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 3: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.35}_{\text{(C) refuse}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 4: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.35}_{\text{(D) refuse}} = 0.0961$

The probability of exactly one 1 of 4 people refusing to administer the shock is the sum of all of these probabilities.

$$0.0961 + 0.0961 + 0.0961 + 0.0961 = 4 \times 0.0961 = 0.3844$$

- Each person in Milgram's experiment can be thought of as a *trial*.
- A person is labeled a *success* if she refuses to administer a severe shock, and *failure* if she administers such shock.
- Since only 35% of people refused to administer a shock, *probability of success* is $p = 0.35$.
- When an individual trial has only two possible outcomes, it is called a *Bernoulli random variable*.

The question from the prior slide asked for the probability of given number of successes, k, in a given number of trials, n, ($k = 1$ success in $n = 4$ trials), and we calculated this probability as

$$\# \textit{ of scenarios} \times P(\textit{single scenario})$$

- $\#$ *of scenarios*: there is a less tedious way to figure this out, we'll get to that shortly...
- $P(\textit{single scenario}) = p^k \, (1 - p)^{(n-k)}$

  probability of success to the power of number of successes, probability of failure to the power of number of failures

The *Binomial distribution* describes the probability of having exactly k successes in n independent Bernouilli trials with probability of success p.

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, $n = 9$ and $k = 2$:

*RR*SSSSSSS
S*RR*SSSSSS
SS*RR*SSSSS
...
SS*R*SS*R*SSS
...
SSSSSSS*RR*

writing out all possible scenarios would be incredibly tedious and prone to errors.

Choose function

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- $k = 1$, $n = 4$: $\binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4\times3\times2\times1}{1\times(3\times2\times1)} = 4$
- $k = 2$, $n = 9$: $\binom{9}{2} = \frac{9!}{2!(9-1)!} = \frac{9\times8\times7!}{2\times1\times7!} = \frac{72}{2} = 36$

Note: You can also use R for these calculations:

```
> choose(9,2)
[1] 36
```

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) There are $n-1$ ways of getting $n-1$ successes in n trials, $\binom{n}{n-1} = n-1$.

Symmetric relation

$$\binom{n}{k} = \binom{n}{n-k}, \text{ for } k = 0, 1, \ldots, n.$$

Binomial probabilities

Let p represent probability of success, $(1-p)$ probability of failure, n the number of independent trials.
X is a random variable representing the number of successes among the n trials.
If

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

then X has binomial distribution $X\sim\mathrm{Bin}(n,p)$.

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

(a) the trials must be independent

(b) the number of trials, n, must be fixed

(c) each trial outcome must be classified as a *success* or a *failure*

(d) the number of desired successes, k, must be greater than the number of trials

(e) the probability of success, p, must be the same for each trial

A 2012 Gallup survey suggests that 26.2% of Americans are obese.
Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, $100 \times 0.262 = 26.2$.

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

- Mathematical proof of $\mu = np$.

- Going back to the obesity rate:

$$\mu = np = 100 \times 0.262 = 26.2$$

But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value to vary?

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

- We would expect 26.2 out of 100 randomly sampled Americans to be obese, with a standard deviation of 4.4.

Note: Mean and standard deviation of a binomial might not always be whole numbers, and that is alright, these values represent what we would expect to see on average.

Using the notion that *observations that are more than 2 standard deviations away from the mean are considered unusual* and the mean and the standard deviation we just computed, we can calculate a range for the plausible number of obese Americans in random samples of 100.

$$26.2 \pm (2 \times 4.4) = (17.4, 35)$$

Milgram experiment (cont.)

Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict a severe shock. What is the probability that she stops after the first person?

$$P(1^{st} \text{ person refuses}) = 0.35$$

... the third person?

$$P(1^{st} \text{ and } 2^{nd} \text{ shock}, 3^{rd} \text{ refuses}) = \underset{0.65}{\underline{S}} \times \underset{0.65}{\underline{S}} \times \underset{0.35}{\underline{R}} = 0.65^2 \times 0.35 \approx 0.15$$

... the tenth person?

Geometric distribution describes the probability distribution of the number of *independent and identically distributed (iid)* Bernoulli trials needed to get one success.

- independence: outcomes of trials don't affect each other
- identical: the probability of success is the same for each trial

Geometric probabilities

If p represents probability of success, $(1 - p)$ represents probability of failure, and n represents number of independent trials

$$P(\textit{success on the } n^{th} \textit{ trial}) = (1 - p)^{n-1} p$$

Mean and standard deviation of geometric distribution

$$\mu = \frac{1}{p} \qquad \sigma = \sqrt{\frac{1-p}{p^2}}$$

Going back to Dr. Smith's experiment: How many people is Dr. Smith expected to test before finding the first one that refuses to administer the shock?

- The expected value is $\mu = \frac{1}{p} = \frac{1}{0.35} = 2.86$
- The standard deviation $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.35}{0.35^2}} = 2.3$
- Dr. Smith is expected to test 2.86 people before finding the first one that refuses to administer the shock, give or take 2.3 people.
- These values only make sense in the context of repeating the experiment many many times.