STAB22 TUT21

Chong Chen
University of Toronto, Scarborough
Department of Computer and Mathematical Sciences

March 29, 2019

1 Random variables

A variable may take many values. The values of a variable from a random sample or randomized experiment are random. The variable is then called a **random variable**.

1.1 Discrete random variables

A variable taking a set of separate values is called a **discrete variable**.
Example: The number of reported auto accidents per day in GTA.

1.2 Continuous random variables

A variable taking values in an infinite continuum is called **continuous variable**.
Example: The life-length of a cell phone.

2 Probability distributions

A **probability distribution** lists all possible outcomes and their probabilities of a random variable.
Example: Let X = number of languages in which a person is fluent. According to Statistics Canada, for residents of Canada X has probability distribution $P(0) = 0.02$, $P(1) = 0.81$, and $P(2) = 0.17$, with negligible probability for higher values of X.

3 Parameters

Like a Normal distribution, a probability distribution has parameters describing center and variability. The *mean* describes the center, the *variance* and the *standard deviation* describes variability.

3.1 Mean

The mean of the probability distribution for a discrete random variable X is

$$\mu = \sum_{i=1}^{n} x_i P\left(x_i\right)$$

3.1.1 Example 1

Suppose the following distribution is known to us,

X	$P(X = x)$
-1	$1/2$
1	$1/2$

$$\mu = \sum_{i=1}^{n} x_i P\left(x_i\right) = (-1) * 1/2 + 1 * 1/2 = 0$$

3.1.2 Example 2

Consider a gambling. Every time you bet \$1, and you have 1% to win \$20, 0.5% to win \$100. Otherwise you lose.
And the mean is:

$$\mu = \sum_{i=1}^{n} x_i P\left(x_i\right) = (-1) * 0.985 + 20 * 0.01 + 100 * 0.005 = -0.285$$

NOTE The interpretation is, if you play this game repeatedly, **on average**, you are going to lose \$0.285 **each time**.

3.2 Variance

The variance of X is denoted by σ^2 and is defined as

$$\sigma^2 = \sum_{i=1}^{n} P\left(x_i\right)\left(x_i - \mu\right)^2$$

3.3 Standard deviation

The **standard deviation** of X is denoted by σ, and is the square root of variance σ^2.

$$\sigma = \sqrt{\sum_{i=1}^{n} P\left(x_i\right)\left(x_i - \mu\right)^2}$$

3.3.1 Example 3

Let's find the standard deviation of Example 1.

$$\sigma = \sqrt{\sum_{i=1}^{n} P\left(x_i\right)\left(x_i - \mu\right)^2} = \sqrt{1/2 * (-1-0)^2 + 1/2 * (1-0)^2} = 1$$

4 Bernoulli distributions

A **Bernoulli random variable** has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by 0.
Example: there is a 70% chance that a student admitted in a graduate program attends the program. We introduce the random variable X such that:

X	$P(X = x)$
1	70%
0	30%

This random variable X is called Bernoulli random variable with parameter $p = 0.7$.
Question: could you find out the mean for this Bernoulli distribution?

5 Binomial distributions

A **Binomial random variable** describes the probability of the number of successes in n independent **Bernoulli trials** with probability of a success p. Two parameters to describe a binomial distribution fully are:

- n (the number of trials)
- p (the probability of success at each trial)

Example: suppose that a university announced that it admitted 8 students for the following year's graduate program in music. There is a 40% chance that an admitted student will decide to accept the offer and attend this university. Admitted students accept the offer independently of each other.
The variable X = "Number of students that accept the offer", has a binomial distribution with parameters $n = 8$, $p = 0.4$.

The probability distribution of a Binomial random variable with parameters n and p. And use the Binomial table to find the following probabilities.
$P(X = 2) = 0.2029$,
$P(X = 5) = 0.1239$.

6 Exercises

6.1 Exercise 1

Suppose 30% of the students failed a course. If we randomly choose 6 students,

(a) What are the value of parameters?
(b) What is the probability that none of them failed?
(c) What is the probability that exactly three of them failed?
(d) What is the probability that **at most** two of them failed?

Answers:
(a) $n = 6$, $p = 0.3$.
(b) 0.1176.
(c) 0.1852.
(d) $0.1176 + 0.3025 + 0.3241 = 0.7442$.

6.2 Exercise 2

In clinical trials a certain drug has a 10% success rate of curing a known disease. If 15 people are known to have the disease. What is the probability of at least 2 being cured?

(a) 0.5491
(b) 0.01
(c) 0.4509
(d) 0.15
(e) 0.7564

Answer:
(c). $1 - (0.2059 + 0.3432) = 0.4509$