

# Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

## Statistics and their Distributions

The random variables  $X_1, X_2, \dots, X_n$  are said to form a random sample of size  $n$  if the following conditions are satisfied.

1. Every  $X_i$  has the same probability distribution.
2. The random variables  $X_1, X_2, \dots, X_n$  are independent.

### Example

Using the population of graduating RIT students, select a person from the population and observe his or her GPA. Then, with/without replacement (the population size is big and  $n$  is not big compared to the population size) select a student and then observe his or her GPA. Repeat  $n$  times. We consider  $n$  random variables where  $X_1$  is the GPA of the first selected person,  $X_2$  is the GPA of the second selected person, and so on. Then  $X_1, X_2, \dots, X_n$  for a random sample of size  $n$ . Once we finish the  $n^{\text{th}}$  observation, we have  $n$  numbers:

$$x_1, x_2, \dots, x_n$$

We can consider the sample mean, median, variance, fourth spread, etc as functions. Consider the case  $n = 3$ .

$$(2.9, 3.0, 3.8) \rightarrow 3.233\dots$$

$$(2.5, 3.7, 3.9) \rightarrow 3.466\dots$$

$\bar{X}$  is a statistic and random variable that takes a sample as input and outputs a number. Sample mean, sample median, sample variance, et cetera can be viewed as random variables. Each of them is called a statistic. The probability distribution of  $X_1, X_2, \dots, X_n$  is also called population distribution.

## Example

A certain brand of MP3 player comes in three configurations: a model with 2GB of memory, costing \$80, a 4GB model priced at \$100, and an 8GB version with a price tag of \$120. If 20% of all purchasers choose the 2GB model, 30% choose the 4GB model, then the probability distribution of the cost  $X$  of a single randomly selected MP3 player purchase is given by:

$x$	80	100	120
$p(x)$	0.2	0.3	0.5

$$p(80) = p(X = 80) = 0.2$$

$$\mu = 106$$

$$\sigma^2 = 244$$

Suppose on a particular day, only two MP3 players are sold. Let  $X_1$  be the random variable representing the revenue of the first sale, and  $X_2$  be the random variable representing the revenue of the second sale.  $X_1$  and  $X_2$  are independent and their pmf is  $p(x)$ .  $X_1$  and  $X_2$  form a random sample of size 2.

$x_1$	$x_2$	$p(x_1, x_2)$	$\bar{x}$	$s^2$
80	80	0.04	80	0
80	100	0.06	90	200
80	120	$0.2 \times 0.5$	100	800
100	80	0.06	90	200

## The Distribution of the Sample Mean

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . Then:

1.  $E(\bar{X}) = \mu_{\bar{x}} = \mu$

2.  $V(\bar{X}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

Let  $T_o = X_1 + X_2 + X_3 + \dots + X_n$ . (the sample total)

1.  $E(T_o) = \mu_{T_o} = n\mu$

2.  $V(T_o) = \sigma_{T_o}^2 = n\sigma^2$

Each of  $X_1, X_2, \dots, X_n$  has pmf  $p(x)$  or pdf  $f(x)$ .  $X_1, X_2, X_3, \dots, X_n$  are independent.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

In the above proposition,  $p(x)$  or  $f(x)$  can be any function satisfying the following conditions.

- $p(x) \geq 0$  for any  $x$ .
- $f(x) \geq 0$  for any  $x$ .
- $\sum p(x) = 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

$p(x)$  or  $f(x)$  is the pmf or pdf of  $X_i$  and represents the population distribution. Let  $X_1, X_2, \dots, X_n$  be a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Then for any  $n$ ,  $\bar{X}$  is normally distributed with mean  $\mu_{\bar{x}} = \mu$  and variance  $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$  (same as before).  $T_o$  is normally distributed with  $\mu_{T_o} = n\mu$  and  $\sigma_{T_o}^2 = n\sigma^2$ .

## Central Limit Theorem

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Then if  $n$  is sufficiently large,  $\bar{X}$  has approximately normal distribution with  $\mu_{\bar{x}} = \mu$  and  $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ .  $T_o$ : normal with  $\mu_{T_o} = n\mu$  and  $\sigma_{T_o}^2 = n\sigma^2$ .

You can find all my notes at <http://omgimanerd.tech/notes>. If you have any questions, comments, or concerns, please contact me at [alvin@omgimanerd.tech](mailto:alvin@omgimanerd.tech)