Chapter 4: Random Variables and Distributions

Ott & Longnecker Sections: 4.6-4.10

Duzhe Wang

the Department of Statistics, UW-Madison

Part 2

UNIVERSITY OF WISCONSIN–MADISON

What do we study?

We'll continue RV.

What do we study?

Key Concepts: Bernoulli RV, Binomial random process, Binomial RV, Expectation of a RV, Variance of a RV, Standard deviation of a RV

Outline

1. Bernoulli RV
2. Binomial random process
3. Binomial RV
4. Expectation, variance and standard deviation

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

x	$p(x)$
0	$1/2$
1	$1/2$

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

x	$p(x)$
0	$1/2$
1	$1/2$

This RV is discrete because it can only take values 0 and 1, and the pmf is expressed by listing a probability for each possible realization. **Note that the total of the probabilities is 1, and this will be the case for every RV.**

Example continued

Let's discuss a bit of notation here. The pmf is really a shorthand notation. When we write something like, $p(0) = 1/3$, we are really writing, $P(X = 0) = 1/2$ and we should read this in words as, "the probability that the RV X realizes to the value 0 is $1/2$." You should get in the habit of reading the notation in this way.

Bernoulli RV

The RV X in the example is a special RV:

- We call a RV a **Bernoulli** RV if it can only realize to the values 0 or 1. The probability that it realizes to 1 is called π. A Bernoulli RV X can be denoted as $X \sim \textit{Bern}(\pi)$. The symbol "$\sim$" should be read "distributed as."

In the example toss of coin, $\pi = 1/2$, so we can say, $X \sim \textit{Bern}(1/2)$.

Outline

1. Bernoulli RV
2. Binomial random process
3. Binomial RV
4. Expectation, variance and standard deviation

Binomial random process

A **binomial random process** has the following properties:

- The random process consists of n identical Bernoulli variables which we call **trials**.
- In each Bernoulli trial we call an outcome of 1 a **success**, and an outcome of 0 a **failure**.
- The probability of a success on any single trial is the same for every trial, and is denoted π.
- The trials are **independent**, in that the outcome of any trial does not affect the outcome of any other trial.

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

- **One basketball player attempts 10 free throws and the number of successful attempts is recorded.**

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

- **One basketball player attempts 10 free throws and the number of successful attempts is recorded.**

The process does consist of trials(each attempt), and the result has two outcomes, success and failure. However, it is unlikely that the trials are independent. The performance on the previous attempt might affect the probability of making the next attempt. The player might, for example, concentrate harder after a miss. Or they might get discouraged after a miss and concentrate less. Therefore this is not a good approximation to a binomial process.

Binomial random process examples

- **Ten different basketball players each attempt 1 free throw and the total number of success attempts is recorded.**

- **Ten different basketball players each attempt 1 free throw and the total number of success attempts is recorded.**

The process does consist of trials(each attempt), the result has two outcomes, success and failure and as long as the players do not influence one another, they are probably approximate independent. However, it is unlikely that every player has the same chance of making a free throw, therefore the probability of success on each trial is not constant. So this is not a good approximation to a binomial process.

Outline

1. Bernoulli RV
2. Binomial random process
3. Binomial RV
4. Expectation, variance and standard deviation

Binomial RV

If the random process is a binomial random process, we then define:

- A **binomial** RV, call it B, is the total number of successes achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such an RV as $B \sim Bin(n, \pi)$.

Binomial RV

If the random process is a binomial random process, we then define:

- A **binomial** RV, call it B, is the total number of successes achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such an RV as $B \sim Bin(n, \pi)$.

A Binomial RV can be thought of as the sum of n independent Bernoulli RVs that all have the same probability of success π. Thus, the values a $Bin(n, \pi)$ can take are $0, 1, \ldots, n$.

An example of Binomial RV

One practical example that closely approximates a binomial random process is the manufacture of circuit boards. Suppose, based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If we select 5 boards at random and then record the number of operational boards in that sample of 5, this process can be described by a binomial process.

Let $Y =$ the number of operational boards in that sample of 5, then

An example of Binomial RV

One practical example that closely approximates a binomial random process is the manufacture of circuit boards. Suppose, based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If we select 5 boards at random and then record the number of operational boards in that sample of 5, this process can be described by a binomial process.
Let $Y =$ the number of operational boards in that sample of 5, then

$$Y \sim Bin(5, 0.96)$$

Pmf of Binomial RV

For a binomial RV $B \sim Bin(n, \pi)$, the probability of observing b successes is:

$$p(b) = \tfrac{n!}{b!(n-b)!}\pi^b(1-\pi)^{n-b},$$

where $n!$ is called the factorial, and is the product of all integers from n to 1. By definition, $0! = 1$. For the circuit board example, we could calculate the probability that the RV Y realizes to 4 (i.e. the chance 4 of the 5 boards are working) as:

$$p(4) = P(Y = 4) = \tfrac{5!}{4!(5-4)!}0.96^4(1-0.96)^{5-4} = 0.17.$$

Outline

1. Bernoulli RV
2. Binomial random process
3. Binomial RV
4. Expectation, variance and standard deviation

Expectation of a RV

The following are important properties of random variables we will see many times in this course.

The **expectation** or **expected value** of a RV X, denoted $E(X)$ or μ_X, is the mean of the population. The expectation of a *discrete* RV X is:

$$\mu_X = E(X) = \sum_x x * p(x)$$

where the sum is taken over all possible realizations of X. (Note that we can define the expected value for continuous RVs using the pdf and *integrals*, which allow you to take continuous sums.)

Variance of a RV

The **variance** of a RV X, denoted $VAR(X)$, or σ_X^2 is the variance of the population. The variance of a *discrete* RV X is:

$$\sigma_X^2 = VAR(X) = \sum_x p(x) * (x - E(X))^2.$$

(Once again, if we were dealing with the variance of a continuous RV, we'd replace the pmf with the pdf and the sum with an integral.)

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

$$E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$$

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

$$E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$$
$$VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$$

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

$$E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$$
$$VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$$
$$SD(X) = \sqrt{VAR(x)} = 1/2$$

Standard deviation of a RV

We could equally well talk about standard deviation of a RV, where $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

$$E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$$
$$VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$$
$$SD(X) = \sqrt{VAR(x)} = 1/2$$

What do you find?

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1 - \pi)$. This isn't hard to prove, but we skip the details.

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1 - \pi)$. This isn't hard to prove, but we skip the details.

What about for binomial RV?

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1 - \pi)$. This isn't hard to prove, but we skip the details.

What about for binomial RV?

It can be shown that

$$\text{if } B \sim Bin(n, \pi), \text{ then } E(B) = n\pi, \text{ and } VAR(B) = n\pi(1 - \pi).$$

For now you will have to take this on faith, but we will see how to derive these results later.

Example. Back to the circuit boards. Recall that Y denotes the number of operational boards in five random boards, and $Y \sim Bin(5, 0.96)$. Then

$$E(Y) = 5 * 0.96 = 4.8$$
$$Var(Y) = 5 * 0.96 * 0.04 = 0.192$$

What's the next?

In the next lecture we will discuss continuous random variables.