Lecture 7

09/20/19

Discrete Distributions: Bernoulli and Binomial Distributions

Manju M. Johny

STAT 330 - Iowa State University

Discrete Distributions

Discrete Distributions

Common distributions for discrete random variables

- Bernoulli distribution RV "is distributed"

$$X \sim Bern(p)$$ parameter

- Binomial distribution

$$X \sim Bin(n, p)$$ parameters

- Geometric distribution

$$X \sim Geo(p)$$ parameter

- Poisson distribution

$$X \sim Pois(\lambda)$$ parameter

We will also discuss *joint distributions* for 2 or more discrete random variables

Bernoulli Distribution

Bernoulli Distribution

Bernoulli Experiment: Random experiment with only 2 outcomes:

- Success (S)
- Failure (F)

where P(Success) = P(S) = p for $p \in [0, 1]$

Example 1: (Bernoulli experiments):

1. Flip a coin: S = heads, F = tails
2. Watch stock prices: S = increase, F = decrease
3. Cancer screening: S = cancer, F = no cancer

Working with Bernoulli Random Variable

Suppose we have a situation that matches a Bernoulli experiment (only 2 outcomes: "success" and "failure").

We obtain the outcome "success" with probability p

When random variable X follows a *Bernoulli Distribution*, we write

R.V
↓
$$X \sim Bern(p)$$
↙ parameter that my dist depends on
↑ "is distributed"

- Define a random variable X

$$X = \begin{cases} 1 & \text{Success (S)} \\ 0 & \text{Failure (F)} \end{cases}$$

Bernoulli Random Variable Cont.

- Probability Mass Function (pmf)
  1. $Im(X) = \{0, 1\}$
  2. $P(X = 1) = P(S) = p$
     $P(X = 0) = P(F) = 1 - p$

The pmf can be written in tabular form:

x	0	1
$p_X(x)$	$1-p$	p

The pmf can be written as a function:

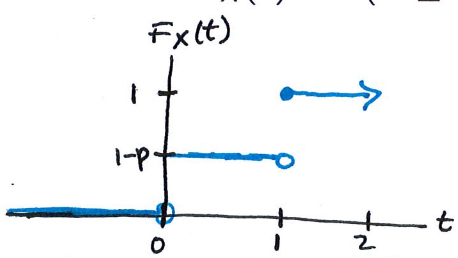
$$p_X(x) = \begin{cases} p^x(1-p)^{1-x} & x \in \{0, 1\} \\ 0 & \text{otherwise} \end{cases}$$

Typically, we use the above functional form to describe the *probability mass function (pmf)* of Bernoulli random variable.

Bernoulli Random Variable Cont.

- Cumulative distribution function (cdf)

$$F_X(t) = P(X \le t) = \begin{cases} 0 & t < 0 \\ 1-p & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases}$$



- Expected Value: $E(X) = p$

$$E(X) = \sum_{x \in \{0,1\}} xP(X = x) = 0(1-p) + 1(p) = p$$

- Variance: $Var(X) = p(1-p)$

$$E(X^2) = \sum_{x=0,1} x^2 P(X=x) = (0^2)(1-p) + (1^2)(p) = p$$

$$Var(X) = E(X^2) - (EX)^2 = p - p^2 = p(1-p)$$

Binomial Distribution

Binomial Distribution

Set up: Conduct multiple trials of *identical* and *independent* Bernoulli experiments

- Each trial is independent of the other trials
- $P(Success) = p$ for each trial

We are interested in the number of success after n trials. The random variable X is

$$X = \text{“ \# of successes in } n \text{ trials”}$$

This random variable X follows a *Binomial Distribution*

$$X \sim Bin(n, p)$$

n, p are parameters that the distribution depends on

where n is the number of trials, and p is the probability of success for each trial.

for different value of n, p, my dist. will look different.

Binomial Distribution Cont.

Example 2: Flip a coin 10 times, and record the number of heads. $n = 10$

Success = "heads"; $P(\text{Success}) = p = 0.5$

- Define the random variable X

$$X = \text{" \# of heads in } n \text{ trials"}$$

10

- The distribution of X is ...

$$X \sim Bin(10, 0.5)$$

$n = 10$, $p = 0.5$

Derivation of Binomial pmf

- Probability Mass Function (pmf) For example 2
  1. $Im(X) = \{0, 1, 2, 3, 4, \ldots, n\}$ $= \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
  2. $P(X = x) = ?$

Recall $P(\text{Success}) = P(S) = p$, $P(\text{Failure}) = P(F) = 1 - p$

Case: $X = 0$ F F F ⋯ F n trials

$P(X = 0) = (1 - p)^n$

Case: $X = 1$ →

S F F F . . . F
F S F F . . . F
F F S F . . . F
⋮
F F F F . . . S

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

Case: $X = 2$

$P(X = 2) = \binom{n}{2} p^2 (1 - p)^{n-2}$

Binomial Random Variables

In general, the *probability mass function (pmf)* of a Binomial R.V can be written as:

$$P(X=x) = p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- Cumulative distribution function (cdf)

$$F_X(t) = P(X \le t) = \sum_{x=0}^{\lfloor t \rfloor} \binom{n}{x} p^x (1-p)^{n-x}$$

$\lfloor t \rfloor$ = "floor" of t

$\lfloor 5 \rfloor = 5$

$\lfloor 4.999\bar{9} \rfloor = 4$

$\lfloor 5.75 \rfloor = 5$

$\lfloor 4.8 \rfloor = 4$

(Add up the pmfs to obtain the cdf)

- Expected Value: $E(X) = np$
- Variance: $Var(X) = np(1-p)$

(independent & identically distributed)

IID Random Variables

Properties of IID Random Variables

Independent and identically distributed (iid) random variables have properties that simplify calculations

Suppose $Y_1, \ldots, Y_n$ are iid random variables

- Since they are *identically* distributed,

$$E(Y_1) = E(Y_2) = \ldots = E(Y_n)$$
$$\rightarrow \underbrace{E(\sum Y_i) = \sum E(Y_i)}_{\text{always}} \underbrace{= nE(Y_1)}_{\text{since } Y_i\text{'s identical}}$$
$$Var(Y_1) = Var(Y_2) = \ldots = Var(Y_n)$$

- Since they are also *independent*,

$$\rightarrow \underbrace{Var(\sum Y_i) = \sum Var(Y_i)}_{\text{since independent (no covariance)}} \underbrace{= nVar(Y_1)}_{\text{since identical}}$$

Working with IID Random Variables

A Binomial random variable, X, is the sum of n *independent and identically distributed (iid)* Bernoulli random variables, Y_i.

Let Y_i be a sequence of iid Bernoulli R.V. For $i = 1, \ldots, n$,

$$Y_i \overset{iid}{\sim} Bern(p)$$

with $E(Y_i) = p$ and $Var(Y_i) = p(1-p)$. Then,

$$X = \sum_{i=1}^{n} Y_i \sim Bin(n, p)$$

Then, we obtain E(X) and Var(X) using properties of iid R.V.s

$$E(X) = nE(Y_1) = np$$
$$Var(X) = nVar(Y_1) = np(1-p)$$

Examples

Binomial Distribution Examples

Example 3: A box contains 15 components that each have a defective rate of 5%. What is the probability that ...

1. exactly 2 out of 15 components are defective?
2. at most 2 components are defective?
3. more than 3 components are defective?
4. more than 1 but less than 4 components are defective?

How should we approach solving these types of problems?

Always start by

1. Defining the random variable
2. Determine the R.V's distribution (and values for the parameters)
3. Use appropriate pmf/cdf/E(X)/Var(X) formulas to solve

Binomial Distribution Examples Cont.

Define the R.V: $X = \#$ of defective components (out of 15 components)

State the Distribution of X: $X \sim Bin(15, 0.05)$ (n ↗, ↖ p)

$n = 15,\ p = 0.05$

1. What is the probability that exactly 2 out of 15 components are defective?

$$P(X = 2) = P_X(2) = \binom{15}{2}(0.05)^2(0.95)^{15-2}$$
$$= \binom{15}{2}(0.05)^2(0.95)^{13}$$
$$= \frac{15!}{2!\,13!}(0.05)^2(0.95)^{13}$$
$$= (105)(0.05)^2(0.95)^{13}$$
$$= 0.1348$$

Binomial Distribution Examples Cont.

2. What is the probability that at most 2 components are defective?

(using pmf) $P(X \leq 2) = P_X(0) + P_X(1) + P_X(2)$

$$= \binom{15}{0}(0.05)^0(0.95)^{15} + \binom{15}{1}(0.05)^1(0.95)^{15-1} + \binom{15}{2}(0.05)^2(0.95)^{15-2}$$
$$= 0.9638$$

(using cdf) $P(X \leq 2) = F_X(2) = 0.9638$ (using Appendix A Binomial Table)

Binomial Distribution Examples Cont.

3. What is the probability that more than 3 components are defective? $P(X > 3) = ? = 1 - P(X \le 3)$

(using pmf) $P(X \le 3) = p_X(0) + p_X(1) + p_X(2) + p_X(3)$

$$= \binom{15}{0}(0.05)^0(0.95)^{15} + \binom{15}{1}(0.05)^1(0.95)^{14} + \binom{15}{2}(0.05)^2(0.95)^{13} + \binom{15}{3}(0.05)^3(0.95)^{12}$$

$$= 0.9945$$

$\Rightarrow P(X > 3) = 1 - P(X \le 3) = 1 - 0.9945 = 0.0055$

(using cdf) $P(X \le 3) = F_X(3) = 0.9945$

$\Rightarrow P(X > 3) = 1 - P(X \le 3) = 1 - 0.9945 = 0.0055$

Appendix A Binomial Table

Binomial Distribution Examples Cont.

4. What is the probability that more than 1 but less than 4 components are defective? $P(1 < X < 4) = ?$

(using pmf) $P(1 < X < 4) = P(X = 2) + P(X = 3)$

$$= p_X(2) + p_X(3)$$

$$= \binom{15}{2}(0.05)^2(0.95)^{13} + \binom{15}{3}(0.05)^3(0.95)^{12}$$

$$= 0.1655$$

(using cdf) $P(1 < X < 4) = P(X < 4) - P(X \le 1)$

$$= P(X \le 3) - P(X \le 1)$$

$$= F_X(3) - F_X(1)$$

$$= 0.9945 - 0.8290$$

$$= 0.1655$$

Appendix A Binomial Table

To use CDF method, we have to write probabilies w/ "$\le$" sign.

CDF table gives $P(X \le t)$