4.2 Bernoulli and Binomial Distributions

Definition 4.2

- A random variable X is defined to have a Bernoulli distribution if the probability function of X is given by
$$f_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1;$$
where the parameter p satisfies $0 < p < 1$.
$f_X(x) = 0$ for other X values.
- $(1-p)$ is often denoted by q.
- $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p = q$.

ST2334 Probability and Statistics CYM Special Probability Distributions 4-11

One can also write the pmf for the Bernoulli distribution as

$$f_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

But it is very important and convenient to use this distribution by writing the pmf in the unified analytical form given in the slide. Make sure you understand how and why it can be written in such a form.

Parameter and Family of Distributions

Remarks:

- Suppose $f_X(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution.
- Such a quantity is called a **parameter** of the distribution.
- p is the **parameter** in the Bernoulli distribution.
- The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

"Parameter" is an important terminology in statistical distributions.

✓ Usually, when we talk about a family of distributions (e.g., bernoulli distribution, binomial distribution, Poisson distribution, and normal distribution), we mean that the pdf/pmf of the distribution is known up to one or several (unknown) parameters.

✓ In many statistical problem, the problem is to make inference on these unknown parameters in the specific distribution. We shall see this in the coming chapters.

Page 4-20

4.2.2 Binomial Distributions

Definition 4.3

- A random variable X is defined to have a **binomial distribution** with two parameters n and p, (i.e. $X \sim B(n,p)$), if the probability function of X is given by
  $$\Pr(X=x) = f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x},$$
  for $x = 0, 1, \cdots, n$, where p satisfies $0 < p < 1$, $q = 1-p$, and n ranges over the positive integers.
- X is the **number of successes** that occur in n **independent Bernoulli trials**.

ST2334 Probability and Statistics CYM Special Probability Distributions 4-20

✓ A random variable X is a $B(n,p)$ random variable IF AND ONLY IF $X = X_1 + X_2 + \ldots + X_n$, where $X_1, \ldots, X_n$ are independent random variables, each of which follows the same Bernoulli distribution with the success probability p. By convention, we say "$X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) $Bernoulli(p)$ random variables".

This is particularly useful when we are to derive some statistical properties of the binomial random variable. For example, To derive the expectation and variance of X, if we use the definitions, it may not be convenient. However if we use the expression $X = X_1 + X_2 + \ldots + X_n$,

★ $E(X) = E(X_1) + E(X_2) + \ldots + E(X_n) = p + p + \ldots + p = np$; and

★ $V(X) = V(X_1 + X_2 + \ldots + V_n) = V(X_1) + V(X_2) + \ldots + V(X_n) = pq + pq + \ldots + pq = npq$, where we have used the property discussed at the end of the complementary notes for week 7.

Note: These results are summarized on page 4-23 of the lecture slides.

✓ In the lecture video for page 4-18, Prof. Chan has already discussed the derivation of the pmf of the binomial distribution given on the lecture slide above. Here we summarize this derivation as follows for your reference.

★ Consider a specific realization of $X_1, \ldots, X_n$, namely $x_1, x_2, \ldots, x_n$ such that $\sum_{i=1}^n x_n = x$. Note the independence of $X_1, X_2, \ldots, X_n$ and that they are all $Bernoulli(p)$ random variables, we have

$$\begin{aligned} Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) &= Pr(X_1 = x_1)Pr(X_2 = x_2)\ldots Pr(X_n = x_n) \\ &= \prod_{i=1}^n p^{x_i} q^{1-x_i} = p^{\sum_{i=1}^n x_i} q^{n-\sum_{i=1}^n x_i} \\ &= p^x q^{n-x}. \end{aligned}$$

★ $\sum_{i=1}^n x_i = x$ on the one hand means that the realized value for the corresponding X is x; on the other hand it means that out of n trials, we get x successes. There are $\binom{n}{x}$ number of such sequences, as we can think of it as choosing x positions to take value 1 out of a length n sequence, and other positions will be 0. As a consequence, by noting that for different choices of $x_1, \ldots, x_n$, $\{X_1 = x_1, X_2 = x_2, \ldots X_n = x_n\}$ are sets of mutually exclusive events, we have

$$\begin{aligned} Pr(X = x) &= Pr\left(\bigcup_{x_1,\ldots,x_n : \sum x_i = x} \{X_1 = x_1, X_2 = x_2, \ldots X_n = x_n\} \right) \\ &= \sum_{x_1,\ldots,x_n : \sum x_i = x} Pr(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) \\ &= \sum_{x_1,\ldots,x_n : \sum x_i = x} p^x q^{n-x} = \binom{n}{x} p^x q^{n-x}. \end{aligned}$$

Page 4-29

Solution to Example 2

- Let X denote the number of children out of the 20 children recover from the disease.
- Then $X \sim B(20, 0.2)$.

(a) $\Pr(X \geq 8) = 0.0321$.

(b) $$\begin{aligned}\Pr(2 \leq X \leq 5) &= \Pr(X \leq 5) - \Pr(X \leq 1) \\ &= 1 - \Pr(X \geq 6) - (1 - \Pr(X \geq 2)) \\ &= \Pr(X \geq 2) - \Pr(X \geq 6) \\ &= 0.9308 - 0.1958 = 0.7350.\end{aligned}$$

Microsoft Excel Worksheet

ST2334 Probability and Statistics CYM Special Probability Distributions 4-29

This is a simple and good review on the contents of Chapter 2: since X is a discrete random variable which may take values of $0, 1, \ldots, 20$, so

$$\begin{aligned}Pr(2 \leq X \leq 5) &= Pr(X \leq 5) - Pr(X < 2) \\ &= Pr(X \leq 5) - Pr(X \leq 1) \\ &= F_X(5) - F_X(1).\end{aligned}$$

Page 4-36 & 4-37

Example 5 (Continued)

- The laboratory technicians must decide whether the data resulting from the experiment supports that claim the $p \leq 0.10$.
- Let X denote the number of units among 20 sampled that need repair, so $X \sim B(20, p)$. (Why?)
- Consider the decision rule:
  - Reject the claim that $p \leq 0.10$ in favour of the conclusion that $p > 0.10$ if $x \geq 5$, (where x is the observed value of X) and
  - consider the claim plausible if $x \leq 4$.

Example 5 (Continued)

- The probability that the claim is rejected when $p = 0.10$ (an incorrect conclusion) is

$$\Pr(X \geq 5 \text{ when } p = 0.10) = 0.0432$$

- The probability that the claim is not rejected when $p = 0.20$ (a different type of incorrect conclusion) is

$$\begin{aligned} &\Pr(X \leq 4 \text{ when } p = 0.20) \\ &= 1 - \Pr(X \geq 5 \text{ when } p = 0.20) \\ &= 1 - 0.3704 = 0.6296. \end{aligned}$$

This is some advance content to the hypothesis testing, which will be covered in more detail and is fundamental in statistical inference. We give some initial discussion below.

Random variable X is the number of units out of 20 that need repair. So, clearly, if the probability p that a unit needs repair is large, X tends to be a large value, and vice versa. Therefore

- ✓ We shall reject $p \leq 0.1$, i.e., a hypothesis that p is small, and in favour of the statement that $p > 0.1$, if we do observe that X is greater than a certain threshold, c say.
- ✓ Likewise, we shall support $p \leq 0.1$, if we observe that X is smaller than the threshold c.

In fact, this comes up with a decision rule for us to conclude whether we are to believe the "hypothesis" $p \leq 0.1$.

Such a decision rule leads to two possible mistakes that we might make, called type I and type II errors; see page 4-37 and watch the corresponding lecture videos for an idea. We can also summarize them in a table as follows

	$p \leq 0.1$ True	$p > 0.1$ True
Conclude $p \leq 0.1$	Correct Decision	Type II error
Reject $p \leq 0.1$	Type I error	Correct Decision

So $Pr(X \geq 5 \text{ when } p = 0.10) = 0.0432$ is the Type I error, but $Pr(X \leq 4 \text{ when } p = 0.20) = 0.6296$ is the Type II error.

Type I and Type II errors always exist when performing a hypothesis testing. Our role is to find a good decision rule (here a reasonable c) that well balances them.

Page 4-57

Mean and Variance of Poisson RV

Theorem 4.4

If X has a **Poisson** distribution with parameter λ, then

$$E(X) = \lambda$$

and

$$V(X) = \lambda.$$

ST2334 Probability and Statistics CYM Special Probability Distributions 4-57

The method for deriving these results in the subsequent pages of the lecture slides is called “density manipulation”. It is very useful in solving many statistical problem. The key is simply: for an arbitrary probability function $f(x)$,

- ✓ if the distribution is discrete, $\sum_{x\in\{x|f(x)>0\}} f(x) = 1$;
- ✓ if the distribution is continuous, $\int_{-\infty}^{\infty} f(x) = 1$.

Read Pages 4-58, 4-59, and 4-60 in the lecture slides, and view the corresponding lecture videos carefully. Then apply the method to derive the results stated on page 4-43.

Negative Binomial Distribution (Continued)

- If $X \sim NB(k, p)$, then it can be shown that

$$E(X) = \frac{k}{p}$$

and

$$Var(X) = \frac{(1-p)k}{p^2}$$

Binomial distribution, Negative Binomial distribution (which accommodates the geometric distribution as a special case), and the Poisson distribution are all founded on the Bernoulli trials. Their corresponding random variables X, however, are defined differently.

- ✓ For binomial distribution, X is defined to be the number of successes out of n trials.
- ✓ For Negative Binomial distribution. X is defined to be the number of trials needed so that we achieve k successes.
- ✓ For Poisson distribution. X is defined to be the number successes in a period of time or in a specific region.