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.

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**.

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

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

★ $\sum_{i=1}^n x_i = x$ 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 $x_1, \dots, x_n$, $\{X_1 = x_1, X_2 = x_2, \dots X_n = x_n\}$ are sets of mutually exclusive events, we have

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

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

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.

Mean and Variance of Poisson RV

Theorem 4.4

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

$$E(X) = \lambda$$

and

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

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.

Example 1

- The average number of robberies in a day is four in a certain big city.
- What is the probability that six robberies occurring in two days?

<u>Solution</u>

- Let X be the number of robberies in two days.
- Then $X \sim P(\lambda)$ where $\lambda = 2 \times 4 = 8$.
- $\Pr(X = 6) = \dfrac{e^{-8}(8)^6}{6!} = 0.1222.$

Microsoft Excel Worksheet

The following properties of the Poisson distribution are useful:

✓ Let X follows the $Poisson(\lambda_1)$ distribution. Let Y follows the $Poisson(\lambda_2)$ distribution. X and Y are independent. Then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$.

This has been applied in this slide. The average number of robberies in a day is four; therefore, if we define $X_1 = \#$ of robberies in day 1; $X_2 = \#$ of robberies in day 2, then $X_1 \sim P(4), X_2 \sim P(4)$. X_1 and X_2 are independent as they are the number of occurrences of events in different days. So we conclude $X = X_1 + X_2 \sim P(4 + 4) = P(8)$.

✓ Let X be the number of occurrences of events in a period of time T; it has the $Poisson(\lambda)$ distribution. If Y is the number of occurrences of events in the period of time tT, then $Y \sim Poisson(t\lambda)$. Note that this can also be used to identify that $X \sim Poisson(8)$ given in the slide.

Example 3

A can company reports that the number of breakdowns per 8 hour shift on its machine-operated assembly line follows a Poisson distribution, with a mean of 1.5.

(a) What is the probability of exactly two breakdowns during the midnight shift?

(b) What is the probability of fewer than two breakdowns during the afternoon shift?

(c) What is the probability that no breakdowns during three consecutive 8-hour shifts?

(Assume the machine operates independently across shifts.)

In the lecture video, Prof. Chan solved the problem jointly using the Poisson distribution and binomial distribution. This problem, on the other hand, can also be solved using the properties in the last page.

✓ Let X_i = the number of breakdowns during the ith 8-hour shift. Then $X_i \sim Poisson(1.5)$.

✓ Let $Y = X_1 + X_2 + X_3$, then Y is the number of breakdowns during three consecutive 8-hour shift. Therefore $Y \sim Poisson(4.5)$.

✓ Part (c) is asking $Pr(Y = 0)$, which is $e^{-4.5}\frac{4.5^0}{0!} = e^{-4.5}$.

Page 4-73 &133

4.5 Poisson Approximation to the Binomial Distribution

Theorem 4.5

- Let X be a **Binomial** random variable with parameters n and p. That is

$$\Pr(X = x) = f_X(x) = {}_nC_r\, p^x q^{n-x}, \text{ where } q = 1 - p.$$

- Suppose that $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains a constant as $n \to \infty$.
- Then X will have approximately a Poisson distribution with parameter np. That is

$$\lim_{\substack{p\to 0\\ n\to\infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$$

Theorem

Theorem

- If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$,
- then as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0,1)$$

The approximation of the binomial distribution by Poisson or Normal distributions is under totally different scenarios.

✓ Poisson distribution is a good approximation to the Binomial(n, p) distribution, when n is large but np is small (so that $n(1-p)$ is large).

✓ Normal distribution is a good approximation to the Binomial(n,p) distribution, when n, np, $n(1-p)$ are all large. Practically, we require $np > 5$ and $n(1-p) > 5$.

Keep in mind that these are just approximations, they couldn't give you the exact value. Roughly speaking, how good the approximation is depends on how the corresponding conditions are satisfied.

Mean and Variance of Cont Uniform RV

Theorem 4.6

If X is uniformly distributed over $[a, b]$, then

$$E(X) = \frac{a+b}{2}, \qquad \text{and} \quad V(X) = \frac{1}{12}(b-a)^2.$$

Proof

$$E(X) = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{b}$$

$$= \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

Comparing this properties with the discrete random variables, we observe that the continuous uniform distribution over an interval has nicer formulae for evaluating the expectation and the variance.

But keep in mind that these formulae are applicable only when that the distribution is defined on a single interval.

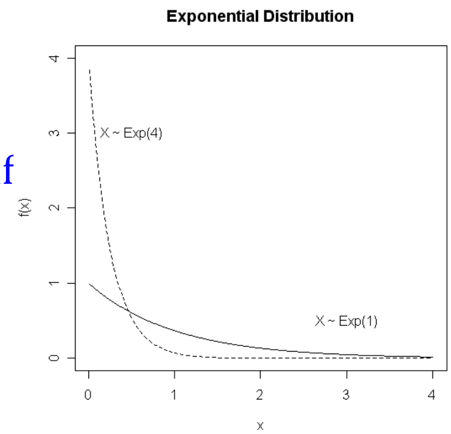
4.7 Exponential Distribution

Definition 4.6

- A continuous random variable X assuming all nonnegative values is said to have an exponential distribution with parameter $\alpha > 0$ if its probability density function is given by

  $f_X(x) = \alpha e^{-\alpha x}$, for $x > 0$.

  and 0 otherwise.
- Note : $\int_{-\infty}^{\infty} f(x)\,dx = 1$



Exponential distribution has two alternative definitions, one is given on the page above; the other is given on page 4-94 that $f_X(x) = \frac{1}{\mu}e^{-x/\mu}$ for $x > 0$, where the parameter $\mu > 0$. Both definitions can be seen frequently in the literature.

✓ These two definitions are equivalent since if we do the reparameterization $\alpha = 1/\mu$.

✓ If we use the definition given on page 4-90, $E(X) = 1/\alpha$ and $V(X) = 1/\alpha^2$; see page 4-91. The cdf is given by $F_X(x) = 1 - e^{-\alpha x}$; see page 4-99.

✓ If we use the definition given on page 4-94, $E(X) = \mu$ and $V(X) = \mu^2$; see page 4-94. The cdf is given by $F_X(x) = 1 - e^{-x/\mu}$.

Exponential distribution is popularly used to model the survival (recovery) time of a patient in the medical research, where $P(X > t) = 1 - F_X(t)$ is called the survival function. It is the probability that the survival (recovery) time of a patient is greater than t.

4.8 Normal Distribution

Definition 4.7

- The random variable X assuming all real values, $-\infty < x < \infty$, has a **normal** (or **Gaussian**) distribution if its probability density function is given by
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,$$
where $-\infty < \mu < \infty$ and $\sigma > 0$.
- It is denoted by $N(\mu, \sigma^2)$.
- μ and σ are called parameters of the normal distribution.

Normal distribution is the most important and popularly used distribution in statistics. Pay extra attention to the properties given on pages 4-106 to 4-110.

✓ The density is symmetric about μ, which is the expectation and the median of the distribution; one direct consequence is $Pr(X \le \mu) = Pr(X \ge \mu) = 0.5$. μ is also called the location parameter, which determines the location of the center of the distribution.

✓ $\sigma^2 = V(X)$ is the shape parameter (also called the dispersion parameter in the literature), which determines the shape of the density function.

✓ No matter what are the values for μ and σ^2, the density is positive for $x \in \mathbb{R}$. It gets closer and closer to (but never touch) 0, when x approaches ∞ or approaches $-\infty$.

✓ The standardization $Z = \frac{X-\mu}{\sigma}$ is very important. The density becomes symmetric about 0. That is for any $z \in \mathbb{R}$, $Pr(Z \le -z) = Pr(Z \ge z)$. $E(Z) = 0$ and $V(Z) = 1$. With this standardization, for $x_1 < x_2$, $Pr(x_1 < X < x_2)$ (with μ and σ^2 being any given values) can always be obtained from the table for Z. See page 4-111 of the lecture slides for more details.

Properties of the normal distribution (Continued)

- The importance of the standardized normal distribution is the fact that it is tabulated.
- Whenever X has distribution $N(\mu, \sigma^2)$, we can always simplify the process of evaluating the values of $\Pr(x_1 < X < x_2)$ by using the transformation $Z = (X - \mu)/\sigma$. Hence $x_1 < X < x_2$ is equivalent to
$$(x_1 - \mu)/\sigma < Z < (x_2 - \mu)/\sigma.$$
- Let $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Then
$$\Pr(x_1 < X < x_2) = \Pr(z_1 < Z < z_2).$$

ST2334 Probability and Statistics CYM Special Probability Distributions 4-111

Read the examples thereafter to understand how all the above properties are jointly applied to solve various problems.

Properties of the normal distribution (Continued)

2. The maximum point occurs at $x = \mu$ and its value is
$$\frac{1}{\sqrt{2\pi}\sigma}$$
3. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
4. The total area under the curve and above the horizontal axis is equal to 1.
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1.$$

Deriving $\int_{-\infty}^{\infty} f_X(x)dx = 1$ is beyond the scope of this module. But if you are interested in it, read the link:

https://en.wikipedia.org/wiki/Gaussian_integral

Example 4

- The breakdown voltage of a randomly chosen diode of a particular type is known to be normally distributed.
- What is the probability that a diode's breakdown voltage is within 1 s.d. of its mean value?

Solution

- This question can be answered without knowing either μ or σ^2, as long as the distribution is known to be normal.
- That is, the answer is the same for **any** normal distribution.

For any normal random variable X, the probability that X is within c s.d. of its mean value is always deterministic, where $c > 0$ is a known constant. In particular, assume $X \sim N(\mu, \sigma^2)$,

$$Pr(\mu - c\sigma < X < \mu + c\sigma) = Pr\left(-c < \frac{X - \mu}{\sigma} < c)\right) = Pr(|Z| < c),$$

which does not depend on μ and σ.

Continuity Correction

Note: In the above calculations, we have made the continuity correction to improve the approximation. In general, we have:

(a) $\Pr(X = k) \approx \Pr(k - ½ < X < k + ½)$.

(b) $\Pr(a \leq X \leq b) \approx \Pr(a - ½ < X < b + ½)$.

$\Pr(a < X \leq b) \approx \Pr(a + ½ < X < b + ½)$.

$\Pr(a \leq X < b) \approx \Pr(a - ½ < X < b - ½)$.

$\Pr(a < X < b) \approx \Pr(a + ½ < X < b - ½)$.

(c) $\Pr(X \leq c) = \Pr(0 \leq X \leq c) \approx \Pr(-½ < X < c + ½)$.

(d) $\Pr(X > c) = \Pr(c < X \leq n) \approx \Pr(c + ½ < X < n + ½)$.

Be aware of the correction rules given in this page, when you use normal distribution to approximate the probabilities based on the binomial random variables.