Chapter 4

Binomial and Poisson Distribution

Yen-Yi Ho

Department of Statistics
University of South Carolina

Outline

- Random Variables
- Binomial Distribution
- Binomial Example (n=3)
- Graphing a Binomial Probability Distribution
- Poisson Distribution
- Poisson Approximation of the Binomial Distribution
- Mean and Variance of a Distribution

Random Variable

Definition: Random Variable

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (only one) numerical value is assigned to each sample point.

Random Variable

Definition: Random Variable

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (only one) numerical value is assigned to each sample point.

e.g. $X =$ the number of heads after flipping a fair coin twice.

$$S = \{HH, HT, TH, TT\}$$

Random Variable

Definition: Random Variable

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (only one) numerical value is assigned to each sample point.

e.g. $X =$ the number of heads after flipping a fair coin twice.

$$S = \{HH, HT, TH, TT\}$$

$$x = 0, 1, 2$$

Random Variable

Definition: Random Variable

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (only one) numerical value is assigned to each sample point.

e.g. $X =$ the number of heads after flipping a fair coin twice.

$$S = \{HH, HT, TH, TT\}$$

$$x = 0, 1, 2$$

By convention, we denote random variables by upper case letters towards the end of the alphabet; e.g., W, X, Y, Z, etc. A possible value of X (i.e., a value in the support) is denoted generically by the lower case version x.

Support of A Random Variable

Definition: Support

The set of all possible values of a random variable is called the **support**.

Two Types of R.V.s

Definition: Discrete

Random variables that can assume a **countable** number of values are called **discrete**.

Two Types of R.V.s

Definition: Discrete

Random variables that can assume a **countable** number of values are called **discrete**.

e.g. $X =$ the $\#$ of students applying to medical schools
support: $0, 1, 2, \ldots$

Two Types of R.V.s

Definition: Discrete

Random variables that can assume a **countable** number of values are called **discrete**.

e.g. $X =$ the $\#$ of students applying to medical schools
support: $0, 1, 2, \ldots$

$Y =$ the $\#$ of customers waiting to be served in a restaurant at a particular time.
support: $0, 1, 2, \ldots$

Two Types of R.V.s (cont'd)

Definition: Continuous

Random variables that can assume values corresponding to **any of the points** contained in an or more intervals are called **continuous**.

Two Types of R.V.s (cont'd)

Definition: Continuous

Random variables that can assume values corresponding to **any of the points** contained in an or more intervals are called **continuous**.

e.g. $W =$ the length of time (in minutes) it takes a student to complete a one-hour exam.
support: $[0, 60]$

Two Types of R.V.s (cont'd)

Definition: Continuous

Random variables that can assume values corresponding to **any of the points** contained in an or more intervals are called **continuous**.

e.g. $W =$ the length of time (in minutes) it takes a student to complete a one-hour exam.
support: $[0, 60]$

$V =$ the weight (in pounds) of a food item bought in a supermarket.

Binomial Distribution

Assumptions:

1. **Bernoulli trial model**

   a. The study or experiment consists of n smaller experiments (trials) each of which has only two possible outcomes,

      e.g. dead, alive
      success, failure
      diseased, not diseased

   b. The outcomes of the trials are independent

   c. The probabilities of the outcomes of the trials remain the same from trial to trial

2. Binomial Distribution (cont'd)

2. The probability of obtaining x "successes" in n Bernoulli trials is

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

where $p =$ probability of a "success" in each independent trial and
$q = 1 - p =$ probability of a "failure"

Note 1: Tables, calculators, and computers are available for calculating these probabilities!

3. Binomial Example ($n = 3$)

Let $n = 3$, $p =$ probability of "success"; then X (the number of "successes") can equal 0, 1, 2 or 3.

# Successes	Samples	$P(X = x)$
3	$+++$	$\binom{3}{3} p^3 q^0 = p^3$
2	$++-$ $+-+$ $-++$	$\binom{3}{2} p^2 q = 3p^2 q$
1	$+--$ $-+-$ $--+$	$\binom{3}{1} p q^2 = 3pq^2$
0	$---$	$\binom{3}{0} p^0 q^3 = q^3$

3. Binomial Example ($n = 3$) (cont'd)

- $P(X \leq 1) = P(X = 0) + P(X = 1)$
- $P(X < 1) = P(X = 0)$
- $P(X > 2) = P(X = 3)$
- $P(1 \leq X \leq 2) = P(X = 1) + P(X = 2)$
- $P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) = 1 - P(X = 0)$

3. Binomial Example ($n = 3$) (cont'd)

The probability that a person suffering from a head cold will obtain relief with a particular drug is 0.9. Three randomly selected sufferers from the cold are given the drug.

- $p = 0.9$
- $q = 1 - p = 0.1$
- $n = 3$
- $2^n = 8$ possible combinations

1	2	3	
S	S	S	ppp
S	S	F	ppq
S	F	S	pqp
F	S	S	qpp
F	F	S	qqp
F	S	F	qpq
S	F	F	pqq
F	F	F	qqq

3. Binomial Example ($n = 3$) (cont'd)

The probability that...

(a) Exactly zero (none) obtain relief:

$$P(X = 0) = q \cdot q \cdot q = q^3 = \binom{3}{0} p^0 q^3 = \frac{3!}{0!\,3!} q^3 = q^3 = (0.1)^3 = 0.001$$

(b) Exactly one:

$$P(X = 1) = qqp + qpq + pqq = \binom{3}{1} p^1 q^2 = \frac{3!}{1!\,2!} pq^2 = 3(0.9)(0.1)^2 = 0.027$$

3. Binomial Example ($n = 3$) (cont'd)

(c) More than one:

$$\begin{aligned}
P(X > 1) &= P(X = 2) + P(X = 3) \\
&= ppq + pqp + qpp + ppp \\
&= \binom{3}{2} p^2 q + \binom{3}{3} p^3 q^0 \\
&= 3p^2 q + p^3 \\
&= 3(0.9)^2(0.1) + (0.9)^3 \\
&= 0.2430 + 0.7290 = 0.9720
\end{aligned}$$

or we can write $\quad P(X > 1) = 1 - P(X \leq 1).$

3. Binomial Example ($n = 3$) (cont'd)

(d) Two or fewer:

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

$$\text{or} \quad = 1 - P(X = 3) = 1 - \binom{3}{3} p^3 q^0 = 1 - 0.7290 = 0.2710$$

(e) Two or three:

$$P(X = 2) + P(X = 3) = P(X \geq 2) = 1 - P(X \leq 1) = P(X > 1)$$

(f) Three:

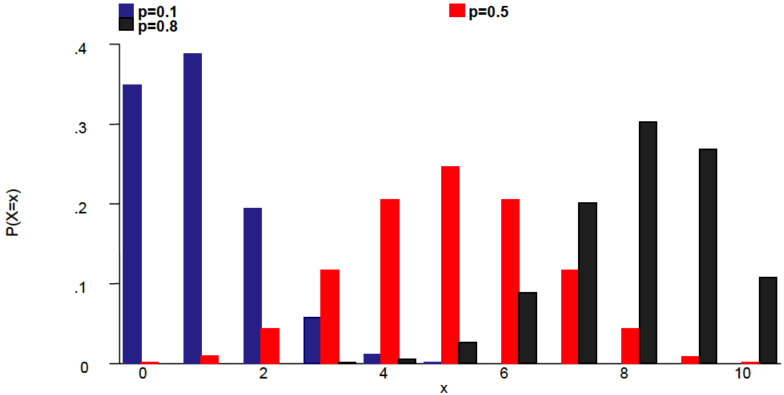
$$P(X = 3) = 0.7290 \quad \text{from above}$$

Binomial Distribution, $n = 10$

		$p = 0.1$	$p = 0.5$	$p = 0.8$
X	X/n	$P(X = x)$	$P(X = x)$	$P(X = x)$
0	0	0.3487	0.0010	0
1	0.1	0.3874	0.0098	0
2	0.2	0.1937	0.0439	0.0001
3	0.3	0.0574	0.1172	0.0008
4	0.4	0.0112	0.2051	0.0055
5	0.5	0.0015	0.2460	0.0264
6	0.6	0.0001	0.2051	0.0881
7	0.7	0	0.1172	0.2013
8	0.8	0	0.0439	0.3020
9	0.9	0	0.0098	0.2684
10	1.0	0	0.0010	0.1074
		1.000	1.000	1.000

5. Graphing a Binomial Probability Distribution

- Center each bar on x; mean $= np$, variance $= npq$. (n=10)
- Total area under histogram equals one.



Changing Histogram with Different Binomial Parameters, n=10

Computing Binomial Probabilities in R

Use R functions `dbinom()` and `pbinom()` to compute probabilities for X:

$$P(X = x) = \texttt{dbinom(x,n,p)}$$

$$P(X \leq x) = \texttt{pbinom(x,n,p)}$$

Example:

Let $X = \#$ free throws you make in 10 attempts. Let $p = 0.7$.

1. Compute $P(X = 3)$.
2. Give the probability that you make at least one free throw.
3. Find $P(X \leq 6)$.
4. Find $P(X > 6)$.

Example: Investment Ventures

Example:

Suppose you invest money into 30 independent ventures. Assume you know the probability that each venture is a success is 0.6.

(a) What is the expected number of ventures that will be a success?

(b) Find the standard deviation of the number of successful ventures.

(a) Expected Number of Successes

For a binomial random variable, the expected value is:

$$\mu = E(X) = np$$

$$\mu = (30)(0.6) = 18$$

Interpretation: On average, we expect 18 of the 30 ventures to be successful.

(b) Standard Deviation

For a binomial random variable, the standard deviation is:

$$\sigma = \sqrt{np(1-p)}$$

$$\sigma = \sqrt{(30)(0.6)(0.4)} = \sqrt{7.2} \approx 2.68$$

6. Poisson Distribution

- Describes the totally random (haphazard) occurrences of events in time or objects in space.
- Often used to describe the distribution of the number of occurrences of a **rare** event.
- Same underlying assumptions as for the binomial distribution — n binary events occur or not at each moment of time (position in space), independently with constant probability, p.

6. Poisson Distribution (cont'd)

- Useful when there are counts with no denominators; λ is known, n and p are unknown.
- Useful when rate and time period are known

Examples of Phenomena that Might Have Poisson Distributions

- Number of Prussian officers killed by horse kicks between 1875 and 1894
- Spatial distribution of stars in a galaxy
- Emergency room visits between 3 am–6 am
- Typographical errors
- Number of deaths due to a rare non-infectious disease

Poisson Probabilities

1. Assumptions

(a) The probability of an event in a short interval is proportional to the length of the interval.

(b) In any extremely small portion of the interval, the probability of more than one occurrence of the event is approximately zero.

(c) Whether or not an event occurs in an interval is independent of events in all other intervals.

Poisson Probabilities (cont'd)

2. The probability of x occurrences of an event in an interval is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

where $\lambda =$ the average or expected number of occurrences of the random event in the interval and $e =$ a constant $(2.718\ldots)$

- For the Poisson distribution, the mean $=$ variance $= \lambda$.

Three Uses of the Poisson Distribution

1. Given $\lambda =$ the expected number of events
2. Given μ and T where μ is a rate (events per time) and $T =$ time period; then $\lambda = \mu T$
3. Given a binomial distribution with n, p; then $\lambda = np$ (Poisson approximation of Binomial)

Poisson Example 1

Example: Bladder Cancer Deaths (Rosner)

Suppose the expected number of deaths due to bladder cancer for all workers in a tire factory is 1.8. If the Poisson distribution is assumed to hold and there are 6 reported deaths due to bladder cancer among the tire workers over the next 20 years, then how unusual is this event?

Poisson probability distribution, $\lambda = 1.8$

$$P(X = 6) = \frac{e^{-1.8}(1.8)^6}{6!} = 0.008$$

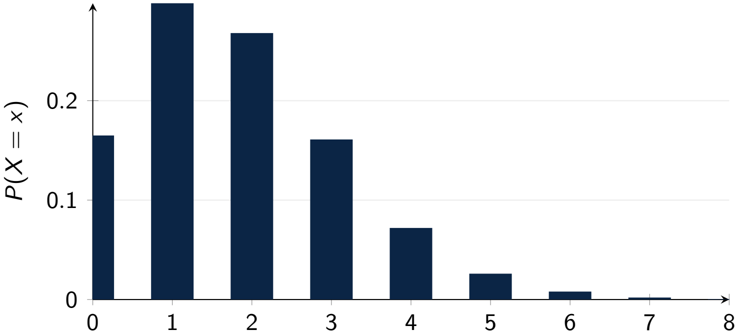
Poisson Example 1 (cont'd)

Poisson probability distribution, $\lambda = 1.8$:

x	$P(X = x)$
0	0.165
1	0.298
2	0.268
3	0.161
4	0.072
5	0.026
6	0.008
7+	0.002
Total	**1.0**

Poisson Example 1 (cont'd)

Poisson probability distribution, $\lambda = 1.8$:



Poisson Example 2

If the rate for a given rare condition is expressed as μ per time period, the expected number of events is $\lambda = \mu T$ where T is the time period.

1. Suppose the weekly rate of suicide in a large city is 2. What is the probability of one suicide in a given week?

Poisson Example 2 (cont'd)

2. What is the probability of 2 suicides in 2 weeks?

Poisson Example 3

Poisson Approximation to the Binomial Distribution

- n is large, p is very small
- or
- $np = \lambda$ is fixed, and n becomes infinitely large
- Example: yearly cases of esophageal cancer in a large city; 30 cases observed in 2000

$$P(X = 6) = \frac{e^{-1.8}(1.8)^6}{6!} = 0.008$$

where $\lambda =$ yearly average number of cases of esophageal cancer

Example — Poisson Approximation of the Binomial Distribution

- Out of 25 babies born to 40-year-old women, what is the frequency of babies with Down's Syndrome?
- Suppose probability of Down's syndrome $= 1/100 = 0.01$; $n = 25$, use binomial distribution
- Expected number of babies with Down's syndrome $= \lambda = np = 25(0.01) = 0.25$
- Use Poisson approximation of the binomial distribution

Example — Poisson Approximation of the Binomial Distribution (cont'd)

	$P(X = x)$	
$x = \#$ babies with Down's Syndrome	Poisson $e^{-\lambda}\lambda^x/x!$	Binomial $\binom{n}{x}p^x q^{n-x}$
0	0.779	0.7778
1	0.195	0.1964
2	0.024	0.0238
> 2	0.002	0.0020
	$\lambda = 0.25$	$n = 25$
		$p = 0.01$

7. Mean and Variance of a Distribution

- Mean of a random variable, X, (the expected value, expectation) is:

$$\mu = E(x) = \begin{cases} \sum_i x_i \cdot P(X = x_i) & \text{for discrete r.v.} \\ \int_{-\infty}^{\infty} x \cdot f(x)\, dx & \text{for continuous r.v.} \end{cases}$$

- Variance of a random variable, X, is:

$$\sigma^2 = \mathrm{Var}(X) = E(X-\mu)^2 = E(X^2) - \mu^2$$

$$\left(\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}(X)} \text{ is the standard deviation}\right)$$

7.1 Mean and Variance — Bernoulli Example

- Let $X = 1$ with probability $= p$; $X = 0$ with probability $= 1 - p$.
- **Calculation of Mean:**

$$E[X] = \mu = \sum_i x_i \cdot P(X = x_i) = 1 \cdot p + 0 \cdot (1 - p) = p$$

- **Variance Calculation:**

$$E[X^2] = (1^2) \cdot p + (0^2) \cdot (1 - p) = p$$

$$\mathrm{Var}[X] = E[X^2] - \mu^2 = p - (p^2) = p \cdot (1 - p)$$

7.2 Properties of Mean and Variance

Some properties of expectation (mean)

- $E(c) = c$ where c is a constant
- $E(cX) = cE(X)$
- $E(X_1 + X_2) = E(X_1) + E(X_2)$
- $E(X_1 - X_2) = E(X_1) - E(X_2)$

Some properties of variance

- $\text{Var}(c) = 0$ where c is a constant
- $\text{Var}(cX) = c^2\text{Var}(X)$
- $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$ only when X_1 and X_2 are independent

The Binomial Distribution

- $S = \sum_{i=1}^{n} X_i$ where X_i are independent Bernoulli(p) random variables
- Thus, S is Binomial(n, p)
- $E[S] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$
- $\text{Var}[S] = \sum_{i=1}^{n} \text{Var}[X_i] = \sum_{i=1}^{n} p(1-p) = np(1-p)$