Statistics I – Chapter 5
Discrete Probability Distributions

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 3, 2012

Introduction

- We have studied **frequency distributions**.
  - For each value or interval, what is the frequency?
- In the next three chapters, we will study **probability distributions**.
  - For each value or interval, what is the probability?
- There are two types of probability distribution:
  - Population distributions: Chapters 5 and 6.
  - Sampling distributions: Chapter 7.

Road map

- **Random variables**.
  - Basic concepts.
  - Expectations and variances.
- Binomial distributions.
- Hypergeometric distributions.
- Poisson distributions.

Random variables

- A **random variable** (RV) is a variable whose outcomes are random.
- Examples:
  - The outcome of tossing a coin.
  - The outcome of rolling a dice.
  - The number of people preferring Pepsi to Coke in a group of 25 people.
  - The number of consumers entering a bookstore at 7-8pm.
  - The temperature of this classroom at tomorrow noon.
  - The average studying hours of a group of 10 students.

Discrete random variables

- A random variable can be discrete, continuous, or mixed.
- A random variable is **discrete** if the set of all possible values is **finite** or **countably infinite**.
  - The outcome of tossing a coin: Finite.
  - The outcome of rolling a dice: Finite.
  - The number of people preferring Pepsi to Coke in a group of 25 people: Finite.
  - The number of consumers entering a bookstore at 7-8pm: Countably infinite.

Continuous random variables

- A random variable is **continuous** if the set of all possible values is uncountable.
  - The temperature of this classroom at tomorrow noon.
  - The average studying hours of a group of 10 students.
  - The interarrival time between two consumers.
  - The GDP per capita of Taiwan in 2013.

Discrete v.s. continuous RVs

- For a discrete RV, typically things are **counted**.
  - Typically there are **gaps** among possible values.
- For a continuous RV, typically things are **measured**.
  - Typically possible values form an **interval**.
  - Such an interval may have a infinite length.
- Sometimes a random variable is called **mixed**.
  - On Saturday I may or may not go to school. If I go, I need at least one hour for communication. Let X be the number of hours I spend in working including communication on Saturday. Then $X \in \{0\} \cup [1, 24]$.
  - By definition, is a mixed RV discrete or continuous?

Discrete and continuous distributions

- The possibilities of outcomes of a random variable are summarized by **probability distributions**, or simply distributions.
- As variables can be either discrete or continuous, distributions may also be either discrete or continuous.
- In this chapter we study discrete distributions.
- In Chapter 6 we study continuous distributions.

Describing a discrete distribution

- On way to fully describe a discrete distribution is to **list all** possible outcomes and their probabilities.
  - Let X be the result of tossing a fair coin:

x	H	T
$\Pr(X = x)$	$\frac{1}{2}$	$\frac{1}{2}$

  - Let X be the result of rolling a fair dice:

x	1	2	3	4	5	6
$\Pr(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Describing a discrete distribution

- But complete enumeration is unsatisfactory if there are too many (or even infinite) possible values.
- Also, sometimes there is a **formula** for the probabilities.
- Suppose we toss a fair coin and will stop with a tail.
- Let X be the number of tosses we make.
  - $\Pr(X = 1) = \frac{1}{2}$ (getting a tail at the first time).
  - $\Pr(X = 2) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$ (head and then a tail).
  - $\Pr(X = 3) = (\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = \frac{1}{8}$ (head, head, and then a tail).
  - In general, $\Pr(X = x) = (\frac{1}{2})^x$ for all $x = 1, 2, ...$
  - No need to create a table!

Probability mass functions

- The formula of calculating the probability of each possible value of a discrete random variable is call a **probability mass function** (pmf).
  - This is sometimes abbreviated as a probability function (pf).
  - $\Pr(X = x) = (\frac{1}{2})^x$, $x = 1, 2, ...$, is the pmf of X.
  - If the meaning is clear, $\Pr(X = x)$ is abbreviated as $\Pr(x)$.
  - Any finite list of probabilities can be described by a pmf.
- In practice, many random variables cannot be exactly described by a pmf (or the pmf is too hard to be found).
- In this case, people may **approximate** the distribution of the random variable by a distribution with a known pmf.
- So the first step is to study some well-known distributions.

Parameters of a distribution

- A distribution depends on a formula.
- A formula depends on some **parameters**.
  - Suppose the coin now generates a head with probability p.
  - How to modify the original pmf $\Pr(X = x) = (\frac{1}{2})^x$?
  - The pmf becomes $\Pr(X = x|p) = p^{x-1}(1 - p)$, $x = 1, 2, ...$
  - The probability p is called the **parameter** of this distribution.
- Be aware of the difference between:
  - The parameter of a population and
  - The parameter of a distribution.

Descriptive measures

- Consider a discrete random variable X with a sample space S, realizations $\{x_i\}_{i \in S}$, and a pmf $\Pr(\cdot)$.
- The **expected value** (or mean) of X is

$$\mu \equiv \mathbb{E}[X] = \sum_{i \in S} x_i \Pr(x_i).$$

- The **variance** of X is

$$\sigma^2 \equiv \mathrm{Var}(X) \equiv \mathbb{E}\left[(X-\mu)^2\right] = \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i).$$

- The **standard deviation** of X is $\sigma \equiv \sqrt{\sigma^2}$.

Descriptive measures: an example

- Let X be the outcome of rolling a dice, then the pmf is $\Pr(x) = \frac{1}{6}$ for all $x = 1, 2, ..., 6$.
  - The expected value of X is

$$\mathbb{E}[X] \equiv \sum_{i=1}^{6} x_i \Pr(x_i) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5.$$

  - The variance of X is

$$\begin{aligned} \operatorname{Var}(X) &\equiv \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i) \\ &= \frac{1}{6}\left[(-2.5)^2 + (-1.5)^2 + \cdots + 2.5^2\right] \approx 2.92. \end{aligned}$$

  - The standard deviation of X is $\sqrt{2.92} \approx 1.71$.

Linear functions of a random variable

- Consider the **linear function** $a + bX$ of a RV X.

Proposition 1

Let X be a random variable and a and b be two known constants, then

$$\mathbb{E}[a + bX] = a + b\mathbb{E}[X] \quad \textit{and} \quad \mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X).$$

Proof. Similar to Problems 5a and 5b in Homework 3.

- If one earns $5x$ by rolling x, the expected value of variance of the earning of rolling a dice are 17.5 and 72.92.

Expectation of a sum of RVs

- Consider the sum of a set of n random variables:

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n.$$

  What is the expectation?

- "Expectation of a sum is the sum of expectations:"

Proposition 2

Let $\{X_i\}_{i=1,\ldots,n}$ be a set of random variables, then

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

Expectation of a sum of RVs

- *Proof of Proposition 2.* Suppose $n = 2$ and S_i is the sample space of X_i, then

$$
\begin{aligned}
\mathbb{E}[X_1 + X_2] &= \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} (x_1 + x_2) \Pr(x_1, x_2) \\
&= \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} x_1 \Pr(x_1, x_2) + \sum_{x_2 \in S_1} \sum_{x_1 \in S_2} x_2 \Pr(x_1, x_2) \\
&= \sum_{x_1 \in S_1} x_1 \sum_{x_2 \in S_2} \Pr(x_1, x_2) + \sum_{x_2 \in S_2} x_2 \sum_{x_1 \in S_1} \Pr(x_1, x_2) \\
&= \sum_{x_1 \in S_1} x_1 \Pr(x_1) + \sum_{x_2 \in S_2} x_2 \Pr(x_2) = \mathbb{E}[X_1] + \mathbb{E}[X_2],
\end{aligned}
$$

where $\Pr(x_1, x_2)$ is the abbreviation of $\Pr(X_1 = x_1, X_2 = x_2)$. $\square$

Expectation of a product of RVs

- Consider the product of n **independent** random variables:

$$\prod_{i=1}^{n} X_i = X_1 \times X_2 \times \cdots \times X_n.$$

Proposition 3

Let $\{X_i\}_{i=1,\dots,n}$ be a set of independent RVs, then

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

Proof. Homework! □

Variance of sum of RVs

- "Variance of an independent sum is the sum of variances:"

Proposition 4

Let $\{X_i\}_{i=1,\dots,n}$ be a set of independent random variables, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

- Is $\mathrm{Var}(2X) = 2\mathrm{Var}(X)$? Why?
- Is $\mathbb{E}(2X) = 2\mathbb{E}(X)$? Why?

Variance of sum of RVs

- *Proof of Proposition 4.* Suppose $n = 2$ and $\mathbb{E}[X_i] = \mu_i$, then

$$\begin{aligned}\mathrm{Var}(X_1 + X_2) &= \mathbb{E}\Big[X_1 + X_2 - \mathbb{E}[X_1 + X_2]\Big]^2 \\ &= \mathbb{E}[X_1 + X_2 - \mu_1 + \mu_2]^2 \\ &= \mathbb{E}\Big[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)\Big] \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathbb{E}\Big[(X_1 - \mu_1)(X_2 - \mu_2)\Big].\end{aligned}$$

Because X_1 and X_2 are independent, $\mathbb{E}[X_1 X_2] = \mu_1 \mu_2$. Thus,

$$\mathbb{E}\Big[(X_1 - \mu_1)(X_2 - \mu_2)\Big] = \mathbb{E}[X_1 X_2] - \mu_1 \mathbb{E}[X_2] - \mu_2 \mathbb{E}[X_1] + \mu_1 \mu_2 = 0,$$

which completes the proof. □

Summary

- Two definitions:
  - $\mathbb{E}[X]$.
  - $\mathrm{Var}(X) = \mathbb{E}\Big[X - \mathbb{E}[X]\Big]^2$.
- Four fundamental properties:
  - $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$ and $\mathrm{Var}[a + bX] = b^2\mathrm{Var}[X]$.
  - $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$.
  - $\mathbb{E}[X_1 \times \cdots \times X_n] = \mathbb{E}[X_1] \times \cdots \times \mathbb{E}[X_n]$ if independent.
  - $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$ if independent.

Road map

- Random variables.
- **Binomial distributions**.
  - Bernoulli distributions.
  - Binomial distributions.
- Hypergeometric distributions.
- Poisson distributions.

Bernoulli trials

- The study of the binomial distribution must start from studying **Bernoulli trials**.
- In some types of trial, the random result is **binary**.
  - Tossing a coin.
  - The sex of a person.
  - Taller or shorter than 170cm.
- One such trial is called a Bernoulli trial.
- This is named after Jacob Bernoulli, the uncle of Daniel Bernoulli, who established the Bernoulli Principle in for fluid dynamics.

Bernoulli distributions

- So in a Bernoulli trial, the outcome is binary.
- Typically they are labeled as **0 and 1**.
  - In some cases, 0 means a failure and 1 means a success.
- Let the **probability of observing 1** be p. This defines the **Bernoulli distribution**:

Definition 1 (Bernoulli distribution)

A random variable X follows the Bernoulli distribution with parameter $p \in (0,1)$, denoted by $X \sim \mathrm{Ber}(p)$, if its pmf is

$$\Pr(x|p) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}.$$

Bernoulli distributions

- What are the mean and variance of a Bernoulli RV?

Proposition 5

Let $X \sim \text{Ber}(p)$, *then* $\mathbb{E}[X] = p$ *and* $\text{Var}(X) = p(1-p)$.

- Intuitions:
  - We will see 1 more likely if p goes up.
  - The variance is zero if $p = 1$ or $p = 0$. Why?
  - The variance is maximized at $p = \frac{1}{2}$. It is the hardest case for predicting the result.

Bernoulli distributions

- *Proof of Proposition 5.* For the mean, we have

$$\mathbb{E}[X] \equiv \sum_{i \in S} x_i \Pr(x_i) = 1 \times p + 0 \times (1-p) = p.$$

For the variance, we have

$$\begin{aligned} \mathrm{Var}(X) &\equiv \sum_{i \in S} (x_i - \mathbb{E}[X])^2 \Pr(x_i) \\ &= (1-p)^2 p + (-p)^2 (1-p) = p(1-p). \end{aligned}$$

Note that both derivations are based on the definitions. $\square$

Some remarks for Jacob Bernoulli

- Jacob Bernoulli (1654 – 1705) was one of the many prominent Swiss mathematicians in the Bernoulli family.
- He is best known for the work *Ars Conjectandi* (The Art of Conjecture), published eight years after his death.
- He discovered **the value of e** by solving the limit

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

- He provided the first rigorous proof for the **Law of Large Numbers** (for the special case of binary variables).

A sequence of Bernoulli trials

- Now we are ready to study the binomial distribution.
- Consider a **sequence** of n **independent** Bernoulli trials.
- Let the outcomes be X_is, where $X_i \sim \mathrm{Ber}(p)$, $i = 1, 2, ..., n$.
- Then consider the **sum** of these Bernoulli variables

  $$Y = \sum_{i=1}^{n} X_i.$$

  Y denotes the number of "1" observed in the n trials.
  - Number of heads observed after tossing a coin ten times.
  - Number of men sampled in 1000 randomly selected people.

Finding the probability: a special case

- What is the probability that we see x 1s in n trials?
- Maybe an easier question: What is the probability that we see two 1s in five trials?
  - There are many different possibilities to see two 1s:

1	1	0	0	0	0	1	1	0	0	0	0	1	1	0
1	0	1	0	0	0	1	0	1	0	0	0	1	0	1
1	0	0	1	0	0	1	0	0	1	0	0	0	1	1
1	0	0	0	1										

  - Note that these are ten mutually exclusive events. What we want is a **union** probability of the union of these ten events.
  - By the **special law of addition**, the union probability is the sum of the probabilities of these ten events.
  - So what is the probability of each event?

Finding the probability: a special case

- The ten events:

1	1	0	0	0	0	1	1	0	0	0	0	1	1	0
1	0	1	0	0	0	1	0	1	0	0	0	1	0	1
1	0	0	1	0	0	1	0	0	1	0	0	0	1	1
1	0	0	0	1										

  - Event 1: $(X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 0, 0)$. This is a **joint** event, an intersection of five **independent** events.
  - So by the **special law of multiplication**, the joint probability is the product of the five marginal events:

$$
\begin{aligned}
& \Pr(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0) \\
= \; & \Pr(X_1 = 1)(X_2 = 1)(X_3 = 0)(X_4 = 0)(X_5 = 0) \\
= \; & p \cdot p \cdot (1 - p)(1 - p)(1 - p) = p^2(1 - p)^3
\end{aligned}
$$

Finding the probability: a special case

- The ten events:

1	1	0	0	0	0	1	1	0	0	0	0	1	1	0
1	0	1	0	0	0	1	0	1	0	0	0	1	0	1
1	0	0	1	0	0	1	0	0	1	0	0	0	1	1
1	0	0	0	1										

  - So the probability of event 1 is $p^2(1-p)^3$. How about event 2?
  - The probability of event 2 is $p(1-p)p(1-p)(1-p)$, which is also $p^2(1-p)^3$!
  - In fact, the probabilities of all the ten events are all $p^2(1-p)^3$.
- Combining all the discussions above, we have

$$\Pr\left(\sum_{i=1}^{n} X_i = 2 \middle| n = 5, p\right) = 10p^2(1-p)^3.$$

Finding the probability

- What is the probability that we see x 1s in n trials?
  - In n trials, we need to see x 1s and $n - x$ 0s.
  - The probability that those "chosen" trials **all** result in 1 is p^x.
  - The probability that other trials all result in 0 is $(1 - p)^{n-x}$.
  - How many different ways to **choose** x trials out of n trials?

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

  - The product of these three yields the desired probability, as shown in the next page.

Binomial distributions

- The variable $\sum_{i=1}^{n} X_i$ follows the Binomial distribution.

Definition 2 (Binomial distribution)

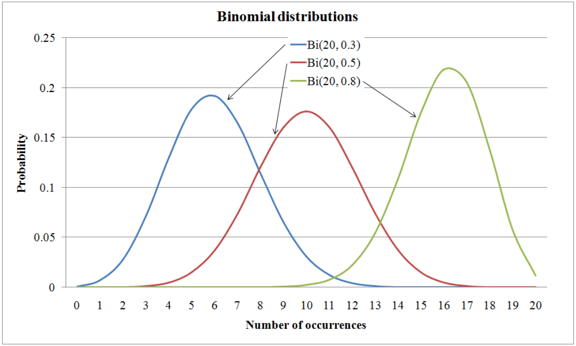
A random variable X follows the Binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$, denoted by $X \sim \mathrm{Bi}(n,p)$, if its pmf is

$$\Pr(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

for $x \in S = \{0, 1, ..., n\}$.

Graphing binomial distributions

- When n is fixed, increasing p shifts the peak of a binomial distribution to the right.



- What is the skewness when $p = 0.5$?

An example

- Suppose a machine producing chips has a 6% defective rate. A company purchased twenty of these chips.
- Let X be the number of defectives, then $X \sim \mathrm{Bi}(20, 0.06)$.
  1. The probability that none is defective is
  $$\Pr(X = 0) = \binom{20}{0} 0.06^0 0.94^{20},$$
  which is around 0.29.
  2. The probability that no more than two are defective is
  $$\begin{aligned}\Pr(X \leq 2) &= 0.29 + \binom{20}{1} 0.06^1 0.94^{19} + \binom{20}{2} 0.06^2 0.94^{18} \\ &= 0.29 + 0.37 + 0.22 = 0.88.\end{aligned}$$

Other applications

- Suppose when one consumer passes our apple store, the probability that she or he will buy at least one apple is 2%. If 100 consumers passes our apple store per day:
  - How many apples may we sell in expectation?
  - Facing the trade off between lost sales and leftover inventory, how many apples should we prepare to maximize our profit?
- Among all candidates we have interviewed, 20% are outstanding. If we randomly hire ten people, what is the probability that at least three of them are outstanding?

Be careful!

- Look at the following “application” again:
  - Among all candidates we have interviewed, 20% are outstanding. If we randomly hire ten people, what is the probability that at least three of them are outstanding?
- Is there anything wrong?
- If there are only fifteen people interviewed, **selecting ten out of fifteen** is NOT a sequence of Bernoulli trails!
- Why?

Sampling with replacement?

- When we sample **without** replacement, we may not use binomial distributions.
  - Randomly selecting six distinct numbers out of 1, 2, ..., 42.
  - Randomly asking ten students in this class regarding whether they want more homework.
- Fortunately, sampling without replacement can be **approximated** by sampling with replacement when $\frac{n}{N} \to 0$.
- In practice, we require $n \leq 0.05N$ for applying the binomial distribution on sampling without replacement.

Expectations and variances

- What are the expectation and variance of a binomial random variable?

Proposition 6

Let $X \sim \mathrm{Bi}(n, p)$, *then*

$$\mathbb{E}[X] = np \quad \textit{and} \quad \mathrm{Var}(X) = np(1-p).$$

- Any intuition?
  - **Hint.** Consider the underlying Bernoulli sequence.

Expectations and variances

- *Proof of Proposition 6.* We can express the binomial random variable as $X = \sum_{i=1}^{n} X_i$, where $X_i \sim \text{Ber}(p)$. Now, according to Proposition 2, we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np.$$

Moreover, according to Proposition 4, we have

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p),$$

where this result is due to the independence of X_is. $\square$

Sum of independent binomial RVs

- What if we add two binomial random variables together?

Proposition 7

Let $X_1 \sim \text{Bi}(n_1, p_1)$ *and* $X_2 \sim \text{Bi}(n_2, p_2)$. *Suppose* X_1 *and* X_2 *are independent and* $p_1 = p_2$, *then then*

$$X_1 + X_2 \sim \text{Bi}(n_1 + n_2, p).$$

- Intuition: It is the sum of two independent Bernoulli sequences.
- What if $p_1 \neq p_2$?

Road map

- Random variables.
- Binomial distributions.
- **Hypergeometric distributions**.
- Poisson distributions.

Hypergeometric distributions

- Consider an experiment with sampling **without** replacement.
- When $n \leq 0.05N$, we may use a binomial distribution to model the experiment.
- What if $n > 0.05N$?
- The **hypergeometric distribution** is defined for this situation.

Hypergeometric distributions

- In describing an experiment like this, we need **three parameters**:
  - N: the population size.
  - A: the number of outcomes that are labeled as "1."
  - n: the sample size.
- Consider a box containing N balls where A of them are white. Suppose we randomly pick up n balls, what is the probability for us to see x white balls?

Hypergeometric distributions: the pmf

- The pmf of a hypergeometric random variable is "a combination of three combinations:"

Definition 3 (Hypergeometric distribution)

An RV X follows the hypergeometric distribution with parameters $N \in \mathbb{N}$, $n \in \{1, 2, ..., N-1\}$, and $A \in \{0, 1, ..., N\}$, denoted by $X \sim \mathrm{HG}(N, A, n)$, if its pmf is

$$\Pr(x|N, A, n) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}$$

for $x \in S = \{0, 1, ..., n\}$.

Expectations and variances

- What are the expectation and variance of a hypergeometric random variable?

Proposition 8

Let $X \sim \text{HG}(N, A, n)$ *and* $p = \frac{A}{N}$, *then*

$$\mathbb{E}[X] = np \quad \textit{and} \quad \text{Var}(X) = np(1-p)\left(\frac{N-n}{N-1}\right).$$

Proof. Homework! □

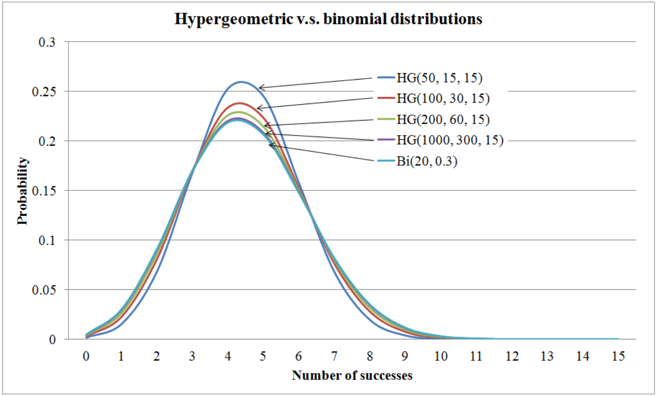
- Similar to those of a binomial random variable?

Expectations and variances

- Consider a binomial RV and a hypergeometric RV:
  - Their means are the same: $np = n\left(\frac{A}{N}\right)$.
  - Their variances are different: $np(1-p)$ and $np(1-p)\left(\frac{N-n}{N-1}\right)$.
- For the two variances, which one is smaller?
- Why? Why sampling with replacement has a **larger** variance than sampling without replacement does?

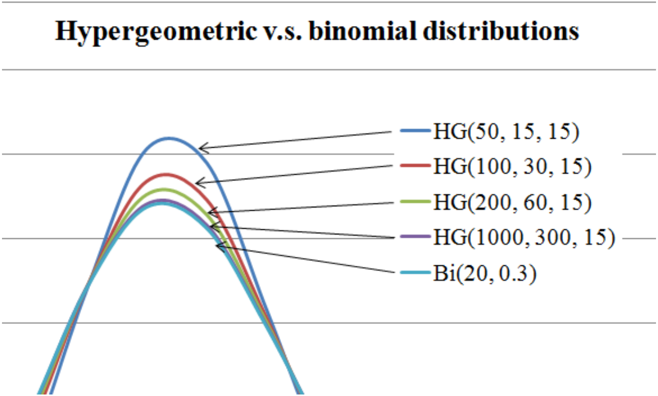
Binomial v.s. hypergeometric RVs

- A hypergeometric random variable can be **approximated** by a binomial random variable **when $\frac{n}{N}$ is close to 0**.



Binomial v.s. hypergeometric RVs

- Also, a hypergeometric RV is more centralized.



Binomial v.s. hypergeometric RVs

- In general, let $\frac{A}{N} = p$, one can show that

$$\frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} \to \binom{n}{x} p^x (1-p)^{1-x} \quad \text{as } N \to \infty.$$

  This shows that a hypergeometric RV is **approximately** a binomial RV **when $\frac{n}{N}$ is close to 0**.
- It is easier to verify that the mean and variance of a hypergeometric RV approach those of a binomial RV:
  - Mean: they are actually the same: $n\left(\frac{A}{N}\right) = np$.
  - Variance: $np(1-p)\left(\frac{N-n}{N-1}\right) \to np(1-p)$ as $N \to \infty$.

Relationships

$$
\begin{tikzcd}
 & \mathrm{Bi}(n,p) \\
\mathrm{Ber}(p) \arrow[ur, "\sum_{i=1}^{n};\ \text{indep.}"] \arrow[dr, "\sum_{i=1}^{n};\ \text{dep.}"'] & \\
 & \mathrm{HG}(N,A,n) \arrow[uu, "{p=\frac{A}{N},\ \frac{n}{N}\to 0}"']
\end{tikzcd}
$$

Road map

- Random variables.
- Binomial distributions.
- Hypergeometric distributions.
- **Poisson distributions**.

Poisson distributions

- The Poisson distribution is one of the most important probability distribution in the field of Operations Research.
- Like the binomial and hypergeometric distributions, it also counts the **number of occurrences** of a particular event.
- However, it does not have a predetermined number of trials. Instead, it counts the number of occurrences within a given **interval or continnum**.
  - Number of consumers entering an LV store in our hour.
  - Number of telephone calls per minute into a call center.
  - Number of typhoons landing Taiwan in one year.
  - Number of sewing flaws per pair of jeans.
  - Number of times that one catches a cold in each year.

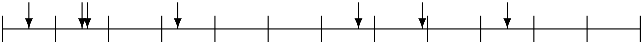
Poisson distributions

- A fundamental assumption of the Poisson distribution is the **homogeneity** of the **arrival rate**.
  - The arrival rate is the rate that the event occurs.
  - The arrival rate is **identical** throughout the interval.
  - It is denoted by λ: In average, there are λ occurrences in one unit of time (be aware of the unit of measurement!).
- Theoretically, the number of occurrence within an interval can range **from zero to infinity**.
- So a Poisson RV can take any nonnegative integer value.
- How to calculate the probability for each possible value?

Poisson distributions: deriving the pmf

- Suppose we want to know the number of occurrences of an event within time interval $[0, 1]$.
  - E.g., number of consumers entering a store in an hour.

- We may divide the interval into n pieces: $[0, \frac{1}{n})$, $[\frac{1}{n}, \frac{2}{n})$, etc.
  - E.g., dividing an hour into twelve 5-minute intervals ($n = 12$).



- We may set n to be large enough so that each piece is short enough and may have **at most one occurrence**.
  - E.g., dividing one hour into 3600 seconds.

Poisson distributions: deriving the pmf

- Each piece is so short that there is at most one occurrence.
  - This can be achieved by making $n \to \infty$.
- Then each piece looks like a **Bernoulli trial** and all pieces are **independent**.
  - For each piece, the probability of one occurrence is $\frac{\lambda}{n}$.
  - Why independent?
- Let X be the number of arrivals in $[0, 1]$ and X_i be the number of arrivals in $[\frac{i-1}{n}, \frac{i}{n})$, $i = 1, ..., n$, then

$$X = \sum_{i=1}^{n} X_i$$

and $X \sim \text{Bi}(n, p = \frac{\lambda}{n})$. Note that $X_i \in \{0, 1\}$.

Poisson distributions: deriving the pmf

- As $X \sim \mathrm{Bi}\big(n, p = \frac{\lambda}{n}\big)$, the pmf is

$$\Pr\left(x|n, p = \frac{\lambda}{n}\right) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \left(\frac{\lambda^x}{x!}\right)\underbrace{\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-x+1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-x}}_{\to 1 \text{ as } n\to\infty!}\left(1-\frac{\lambda}{n}\right)^n.$$

- So $\lim_{n\to\infty}\Pr\left(x\middle|n, p = \frac{\lambda}{n}\right) = \left(\frac{\lambda^x}{x!}\right)\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n.$

Poisson distributions: deriving the pmf

- From elementary Calculus, we have

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

- Therefore,

$$\lim_{n\to\infty}\Pr\left(x\,\middle|\,n, p=\frac{\lambda}{n}\right) = \left(\frac{\lambda^x}{x!}\right)\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \frac{\lambda^x e^{-\lambda}}{x!}.$$

  This is the pmf of a Poisson RV with arrival rate λ.
- A Poisson RV is nothing but the **limiting case** ($n \to \infty$) of a binomial RV!

Poisson distributions: definition

- Now we are ready to define the Poisson distribution.

Definition 4 (Poisson distribution)

A random variable X follows the Poisson distribution with parameters $\lambda > 0$, denoted by $X \sim \text{Poi}(\lambda)$, if its pmf is

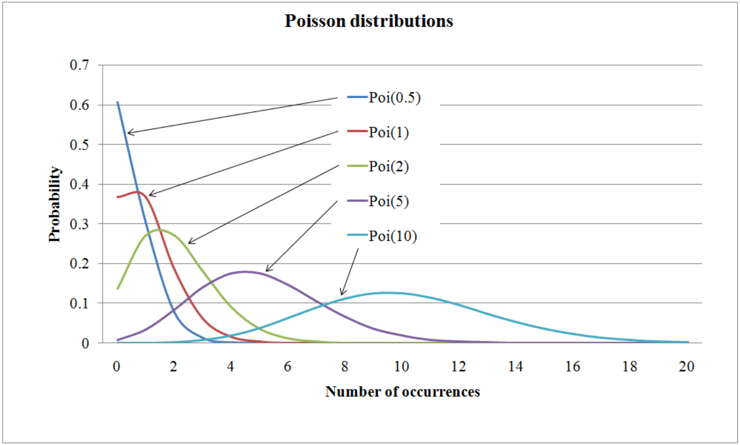
$$\Pr(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for $x \in S = \mathbb{N} \cup \{0\}$.

- It "extends the binomial distribution to infinity."

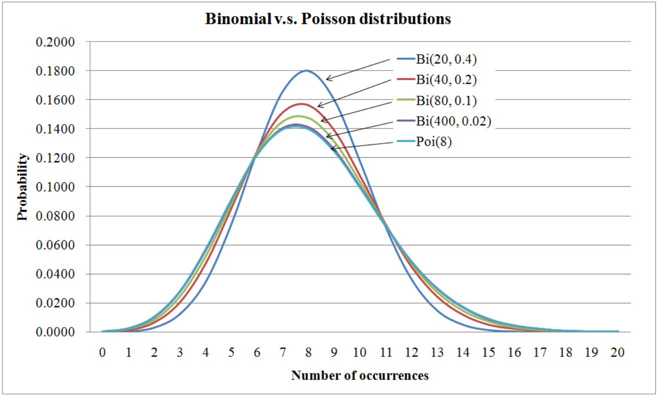
Poisson distributions

- Poisson distributions are **skewed to the right**.



Binomial v.s. Poisson distributions

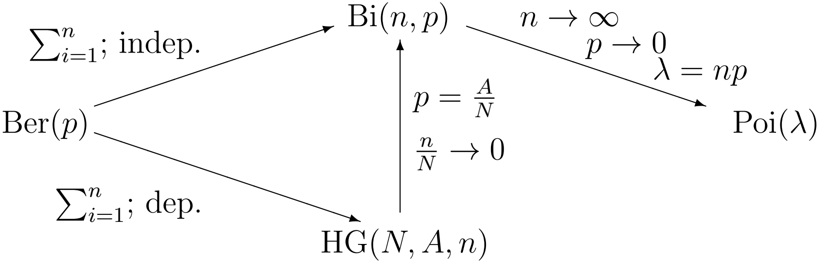
- A Poisson RV can be **approximated** by a binomial RV **when** $n \to \infty$ and $\lambda = np$ **remains constant**.



Binomial v.s. Poisson distributions

- So when n is large and p is small, we may approximate a binomial random variable by a Poisson random variable with $\lambda = np$.
- How large should n be and how small should p be?
- In practice, there are several rule of thumbs:
  - Textbook: when $n \geq 20$ and $np \leq 7$.
  - Dr. Yen: $n > 100$ and $p < 0.01$.
  - Wikipedia: something else.
  - But you know how to verify the quality of approximation.

Relationships



Expectations and variances

- What are the expectation and variance of a Poisson RV?

Proposition 9

Let $X \sim \text{Poi}(\lambda)$, *then*

$$\mathbb{E}[X] = \text{Var}(X) = \lambda.$$

Proof. Later in this semester. □

- Actually, when we say λ is the arrival **rate**, we are implicitly saying that λ is the mean.
- The mean and variance are identical. Is that common?

Time units for Poisson random variables

- Let $X \sim \text{Poi}(\lambda)$. The value of λ depends on the definition of the unit time.
  - If in average 120 consumers enter in one hour, $\lambda = 120/\text{hour}$.
  - Counting in minutes: $\lambda = 2/\text{minute}$.
  - Counting in days: $\lambda = 2880/\text{day}$.
- In short, the value of λ is **proportional** to the length of a unit time.

An example: questions

- The number of car accidents at a particular intersection is believed to follow a Poisson distribution with the mean three per week.
  1. How likely is that there is no accident in one day?
  2. How likely is that there is at least three accidents in a week?
  3. If in the last week there were seven accidents, should you try to reinvestigate the mean of the Poisson distribution?

An example: answers

- Let $X \sim \text{Poi}(3)$ be the number of car accidents at that intersection in one **week**.
  1. Let Y be the number of car accidents at that intersection in one **day**, then $Y \sim \text{Poi}(\frac{3}{7})$. The probability that there is no accident in one day is thus

$$\Pr(Y = 0) = \frac{(\frac{3}{7})^0 e^{-\frac{3}{7}}}{0!} = e^{-\frac{3}{7}} \approx 0.651.$$

An example: answers

- Continued from the previous page:
  2. The probability of at least three accidents in a week is

$$\begin{aligned}\Pr(X \geq 3) &= 1 - \sum_{i=0}^{2} \Pr(X = i) \\ &= 1 - \left(\frac{3^0 e^{-3}}{0!} + \frac{3^1 e^{-3}}{1!} + \frac{3^2 e^{-3}}{2!} \right) \\ &\approx 1 - (0.05 + 0.149 + 0.224) = 0.577.\end{aligned}$$

  3. The probability of seven accidents in a week is

$$\Pr(X = 7) = \frac{3^7 e^{-3}}{7!} \approx 0.022.$$

  It is thus highly possible that λ is larger than we thought.

Summary

- Use **random variables** to model experiments, events, and outcomes.
- Use **distributions** to describe random variables.
- Four important discrete distributions:
  - Bernoulli, binomial, Hypergeometric, and Poisson.
- For each of them, there is a **pmf**, a **mean**, and a **variance**.
- Use them to approximate practical situations and derive probabilities.

Finding the probability

- MS Excel functions.
- The probability tables.
  - Study the textbook by yourself.