Math/Stat 352
Lecture 7

Section 4.1 and 4.2

Commonly Used Distributions: The Bernoulli and Binomial distributions
The Bernoulli Distribution

We use the Bernoulli distribution when we have an experiment/trial which can result in one of two outcomes labeled: Success (S) and Failure (F).

The probability of a success is denoted by \( p = P(S) \).
The probability of a failure is then 1 – p = P(F).

Such a trial/experiment is called a Bernoulli trial with success probability \( p \).
Examples

1. The simplest Bernoulli trial is the toss of a coin. The two outcomes are heads and tails. If we define heads to be the success outcome, then $p$ is the probability that the coin comes up heads. For a fair coin, $p = 0.5$.

2. Another Bernoulli trial is a selection of a component from a population of components, some of which are defective. If we define “success” to be a defective component, then $p$ is the proportion of defective components in the population.
Bernoulli random variable: $X \sim \text{Bernoulli}(p)$

For any Bernoulli trial, we define a random variable $X$ as follows:

If the experiment results in a success, then $X = 1$. Otherwise, $X = 0$.

It follows that $X$ is a discrete random variable, with probability mass function $p(x)$ defined by

$$p(0) = P(X = 0) = 1 - p$$
$$p(1) = P(X = 1) = p$$
$$p(x) = 0 \text{ for any value of } x \text{ other than 0 or 1}$$
Mean and Variance of Bernoulli r.v.

If \( X \sim \text{Bernoulli}(p) \), then

\[
\mu_X = 0(1-p) + 1(p) = p
\]

\[
\sigma^2_X = (0 - p)^2 (1 - p) + (1 - p)^2 (p) = p(1 - p)
\]
Example

Ten percent of components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective, and $X = 0$ otherwise.

1. What is the distribution of $X$?

2. Find the mean and variance of $X$. 
Example

At a certain fast food restaurant, 25% of drink orders are for a small drink, 35% for a medium, and 40% for a large drink. Let $X=1$ if a randomly chosen order is for a small, and let $X=0$ otherwise. Let $Y=1$ if the order is for medium, and $Y=0$ otherwise. Let $Z=1$ if the order is for either small or medium, and let $Z=0$ otherwise.

1. Let $p_x$ denote the success probability for $X$. What is $p_x$?
2. Let $p_y$ denote the success probability for $Y$. What is $p_y$?
3. Let $p_z$ denote the success probability for $Z$. What is $p_z$?
4. Is it possible for both $X$ and $Y$ to equal 1?
5. Does $p_z = p_x + p_y$?
6. Does $Z = X + Y$?
The Binomial Distribution

If a total of \( n \) Bernoulli trials are conducted, and

- The trials are independent.
- Each trial has the same success probability \( p \)
- \( X = \# \) of successes in the \( n \) Bernoulli trials

then \( X \) has the binomial distribution with parameters \( n \) and \( p \), denoted \( X \sim \text{Bin}(n, p) \).
The Binomial Distribution

**Binomial(15, 0.1)**
Unimodal (mode = 1)
Right-skewed
*Mean = np = 1.5)*
Concentrated around 1.5

**Binomial(19, 0.1)**
Unimodal (mode = 1)
Right-skewed
*Mean = np = 1.9*
Concentrated around 1.9
Example

A fair coin is tossed 10 times. Let $X$ be the number of heads that appear. What is the distribution of $X$?
Sampling from Finite Populations: Binomial Distribution

**Experiment:** A simple random sample is drawn from a finite population that contains items of two types: S and F. Assume that the sample size is no more than 5% of the population.

Let $X =$ # of S in the sample. Then, $X \sim \text{Bin}(n, p)$.

**Example:** A lot contains several thousand components, 10% of which are defective. Seven components are sampled from the lot. Let $X$ represent the number of defective components in the sample. What is the distribution of $X$?
Binomial R.V.: pmf, mean, and variance

- If $X \sim \text{Bin}(n, p)$, the probability mass function of $X$ is

$$p(x) = P(X = x) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, & x = 0,1,...,n \\ 0, & \text{otherwise} \end{cases}$$

- Mean of $X$: $EX = \mu_X = np$

- Variance of $X$: $\sigma_X^2 = np(1-p)$
Example

A large industrial firm allows a discount on any invoice that is paid within 30 days. Of all invoices, 10% receive the discount. In a company audit, 12 invoices are sampled at random. What is the probability that fewer than 4 of the 12 sampled invoices receive the discount?
Binomial as sum of iid Bernoulli rv’s.

- Assume $n$ independent Bernoulli trials are conducted.

- Each trial has probability of success $p$.

- Let $Y_1, \ldots, Y_n$ be defined as follows:

$$Y_i = \begin{cases} 1 & \text{if S on ith trial} \\ 0 & \text{if F on ith trial} \end{cases}$$

That is each of the $Y_i$ has the Bernoulli(p) distribution.

- Let $X = \text{number of successes among the } n \text{ trials, so } X \sim \text{Bin}(n, p)$.
  Also note that $X = Y_1 + \ldots + Y_n$.

Thus sum of $n$ iid Bernoulli(p) rv’s has a Binomial(n, p) distribution.
Estimate of $p$

In practice we usually do not know $p$ need to approximate/estimate $p$.

Estimation of $p$: (1) collect a sample on $n$ Bernoulli trials,

(2) compute $X$= the number of $S$ in the sample.

(3) Use the sample proportion of $S$: $\hat{p} = X / n$ to estimate $p$.

Note that: $X \sim \text{Bin}(n, p)$.

Since $X$ is a random variable, then $\hat{p} = X / n$ is also a random variable.
How good is \( \hat{p} \) in estimating \( p \)?

Mean of \( \hat{p} \): \( \mathbb{E}[\hat{p}] = p \), so on average, the estimator \( \hat{p} \) = \( p \).

We say that \( \hat{p} \) is an unbiased estimator of \( p \).

Error of the estimation: \( \hat{p} - p \).

The uncertainty or standard deviation of \( \hat{p} \) is \( \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \).

In practice, when computing \( \sigma \), we substitute \( \hat{p} \) for \( p \), since \( p \) is unknown.
Example

In a sample of 100 newly manufactured automobile tires, 7 are found to have minor flaws on the tread. If four newly manufactured tires are selected at random and installed on a car, estimate the probability that none of the four tires have a flaw.