CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture 3D: The Binomial Distribution

Jimi Oke

UMass Amherst

College of Engineering

October 6, 2025

Outline

1. Introduction
2. The Binomial distribution
3. Mean and variance
4. Outlook

Objectives of today's lecture

Understand and apply the binomial distribution

- PMF
- CDF
- Mean
- Variance

Note about CDF

The median of a distribution is given by the value of X at $F_X(x) = 0.5$.

Example 1: Engineering majors

40% of the students in a university are MIE majors. If four students (A, B, C, D) are chosen at random, what is the probability that exactly one student will be an MIE major?

Solution

In one scenario, student A is an MIE major, while the other 3 are not.

$$\begin{aligned} & P(A = \text{MIE}, B = \text{not}, C = \text{not}, D = \text{not}) \\ = \quad & P(A = \text{MIE}) \times P(B = \text{not}) \times P(C = \text{not}) \times P(D = \text{not}) \\ = \quad & 0.4(0.6)(0.6)(0.6) = 0.4(0.6^3) \\ = \quad & \boxed{0.0864} \end{aligned}$$

But are these all the scenarios?

No. There are 3 others: each of the students B, C or D could also be the MIE major. Thus, the total required probability is $4 \times (0.4)(0.6^3) = 0.346$

Example 1: Engineering majors (cont.)

40% of the students in a university are MIE majors. If four students (A, B, C, D) are chosen at random, now find the probability that exactly two students will be MIE majors?

Solution (cont.)

First, we list the scenarios:

- Scenario 1: $A =$ MIE, $B =$ MIE; $C, D =$ not
- Scenario 2: $A =$ MIE, $C =$ MIE; $B, D =$ not
- Scenario 3: $A =$ MIE, $D =$ MIE; $B, C =$ not
- Scenario 4: $B =$ MIE, $C =$ MIE, $A, D =$ not
- Scenario 5: $B =$ MIE, $D =$ MIE, $A, C =$ not
- Scenario 6: $C =$ MIE, $D =$ MIE, $A, B =$ not

The **number of scenarios** $= \binom{4}{2} = \frac{4!}{2!2!} = 6$.
Each scenario has the same **probability**: $(0.4)(0.4)(0.6)(0.6) = 0.058$
The probability of having 2 MIE majors in a random group of 4 students is:

$$\binom{4}{2}(0.4)^2(0.6)^2 = 6 \times (0.058) = \boxed{0.346}$$

Bernoulli sequence

The Bernoulli sequence describes events which may either occur or *not occur* in N successive trials. Key assumptions:

1. Each trial has only **two possibilities**: occurrence or nonoccurrence
2. The **probability of occurrence** p of the event in each trial is constant
3. The trials are statistically independent

Examples of Bernoulli sequences in engineering

- Operational condition of equipment during a project
- Success or failure of quality control test for manufactured items
- Damage to a building in annual seismic events

The Bernoulli sequence is the basis for the **binomial distribution**

Binomial distribution

Given a Bernoulli sequence with X random number of occurrences of an event, n trials and p the probability of occurrence of each event.

Definition

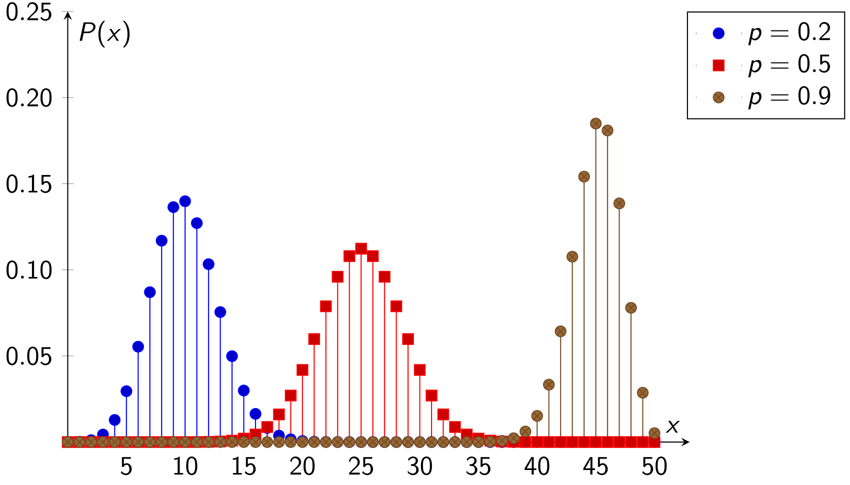
The binomial PMF for a random variable $X \sim \text{Bin}(n, p)$ is given by:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \ldots, n \tag{1}$$

where n and p are the parameters and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient.

- The symbol "$\sim$" is shorthand for "distributed as"
- $\text{Bin}(n, p)$ is the typical notation for a binomial distribution

PMF of a binomial distribution



Example 2: Road graders

Five road graders are used in the construction of a highway project. The probability that a grader will malfunction within 900hrs is 0.0594. Assuming statistical independence among the conditions of the machines, evaluate the probability that two of the five machines will malfunction in less than 900hrs of operation.

Parameters: $n = 5, x = 2, p = 0.0594.$

$$
\begin{aligned}
P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
P(X = 2) &= \binom{5}{2} 0.0594^2 (0.9406)^{5-2} \\
&= 10(0.0035)(0.832) \\
&= \boxed{\mathbf{0.029}}
\end{aligned}
$$

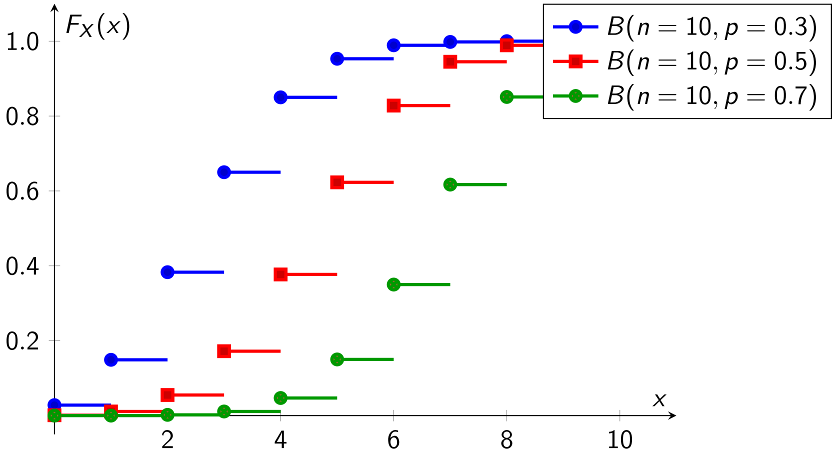
CDF of a binomial distribution

Definition

The CDF of binomially distributed random variable X is:

$$F_X(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \quad (2)$$

CDF of a binomial distribution (visualization)



Example 3: Road graders revisited

5 road graders are used in the construction of a highway project. The probability that a grader will malfunction within 900hrs is 0.0594. Assuming statistical independence among the conditions of the machines, evaluate the prob. **no more than two** of the 5 machines will malfunction within 900hrs of operation.

$$
\begin{aligned}
P(X \leq x) &= F_X(2) = \sum_{k=0}^{2} \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 5, x = 2, p = 0.0594 \\
P(X \leq 2) &= \binom{5}{0} 0.0594^0 (0.9406)^5 + \binom{5}{1} 0.0594^1 (0.9406)^4 \\
&\quad + \binom{5}{2} 0.0594^2 (0.9406)^3 \\
&= (1)(1)(0.9406)^5 + (5)(0.0594)(0.9406)^5 + (10)(0.0594)^2(0.9406)^3 \\
&= \boxed{\mathbf{0.998}}
\end{aligned}
$$

Mean of a binomial distribution

Let $X \sim \text{Bin}(n, p)$:

$$\mu_X = \mathbb{E}(X) = np \tag{3}$$

Proof.

Let $X_i = 1$ if an event occurs on the i-th trial in a Bernoulli sequence. Then the number X of occurrences is: $X = \sum_{i=1}^{n} X_i$.
The expectation is linear in X, thus:

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) \tag{4}$$

Since $\mathbb{E}(X_i) = p$, then:

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = np \tag{5}$$

□

Variance of a binomial distribution

Let $X \sim \text{Bin}(n, p)$. Then

$$\mathbb{V}(X) = np(1-p) = npq \tag{6}$$

where $q = 1 - p$.

Sketch of proof

You can show that variance of a single trial
$\mathbb{V}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = p - p^2 = p(1-p) = pq$. And $\mathbb{V}(X)$ follows.

Example 1: Revisited

40% of the students in a university are engineering majors. The probability that any subset of 4 randomly selected students will be MIE majors is governed by the binomial distribution.

(a) What is the mean of the binomial distribution governing this set of outcomes?

(b) What is the variance?

(c) Find the probability that 2 of 4 randomly selected students will be MIE majors.

(d) Find the probability that at least 3 randomly selected students will be MIE majors.

Example 1: Revisited (cont.)

(a) $n = 4$; $p = 0.4$. Thus, the mean is given by $\mathbb{E}(X) = np = 4(0.4) = \boxed{1.6}$.

(b) The variance is given by $\mathbb{V}(X) = npq$
$= np(1-p) = 4(0.4)(1-0.4) = 4(0.4)(0.6) = \boxed{0.96}$.

(c) Find the probability that 2 of 4 randomly selected students will be MIE majors.

$$\begin{aligned} P(X=2) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{4}{2}(0.4)^2(0.6)^{4-2} \\ &= 6(0.16)(0.36) \\ &= \boxed{0.346} \end{aligned}$$

In Python:

```python
from scipy.stats import binom
n = 4
p = 0.4
x = 2
prob = binom.pmf(x, n, p)
print(prob)
```

Example 1: Revisited (cont.)

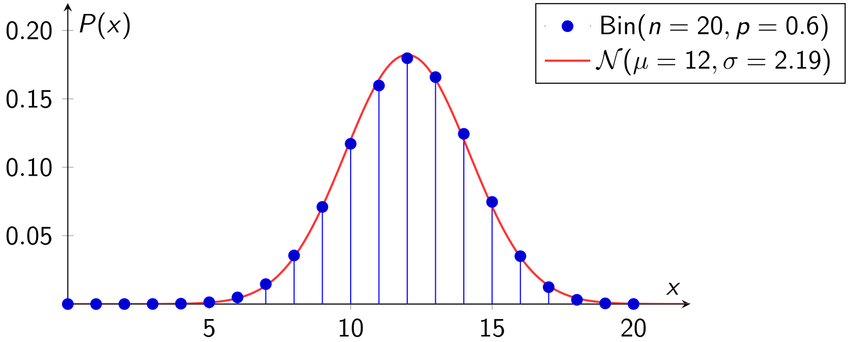
(d) Find the probability that at least 3 randomly selected students will be MIE majors.

We want to find $P(X \geq 3) = 1 - P(X \leq 2) = 1 - F_X(2)$ In Python:

```
from scipy.stats import binom
n = 4
p = 0.4
x = 3
prob = 1 - binom.cdf(x-1, n, p)  # P(X >= 3)
print(prob)
```

Relationship between binomial and normal distributions

Consider the distribution $B(n = 20, p = 0.6)$.
We see that it can be approximated by $\mathcal{N}(\mu = np, \sigma = \sqrt{npq})$, where $q = 1 - p$.



Relationship between binomial and normal (cont.)

If a binomial PMF is not too skewed, then $X \sim \text{Bin}(n, p)$ is approximately normally distributed with $\mu = np$ and $\sigma = \sqrt{npq}$.
To check for normality, we can use the following rules of thumb:

$$np \geq 10 \tag{7}$$

$$nq \geq 10 \tag{8}$$

Thus:

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right) \quad np \geq 10; nq \geq 10 \tag{9}$$

and

$$P(X \geq x) \approx 1 - \Phi\left(\frac{x - 0.5 - np}{\sqrt{npq}}\right) \quad np \geq 10; nq \geq 10 \tag{10}$$

where $\Phi(z)$ is the CDF of the standard normal distribution and ± 0.5 is the **continuity correction**.

Recap: Binomial distribution

- Mean: $\mu_X = E(X) = np$
- Variance: $Var(X) = npq = np(1-p)$
- PMF: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
- CDF: $F_X(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$

Reading

- Open Intro Statistics: Section 4.3 (Binomial distribution)