

Random Variables and Binomial Distribution

Textbook Sections 3.1–3.2

Emile Latour, Nicky Wakim, Meike Niederhausen

January 26, 2026



Learning Objectives

By the end of today's class, you should be able to:

1. Define random variables and distinguish between discrete and continuous random variables
2. Calculate the expected value (mean) and variance of discrete random variables
3. Calculate the expected value and variance of linear combinations of random variables
4. Identify when the Binomial distribution is appropriate and calculate probabilities using it



Roadmap for Today

Part 1: Random Variables (Section 3.1)

- What are random variables?
- Expected value and variance
- Linear combinations of random variables

Part 2: Binomial Distribution (Section 3.2)

- Bernoulli distribution
- Binomial distribution
- Calculating probabilities with the Binomial



Part 1: Random Variables

Section 3.1



Motivation: From Data to Probability Models

So far we've worked with:

- **Data points:** $x_1, x_2, x_3, \dots, x_n$ (observed values)
- **Sample statistics:** mean, variance, standard deviation

Now we're moving to:

- **Random variables:** mathematical models for uncertain outcomes
- **Probability distributions:** theoretical descriptions of random phenomena
- **Parameters:** population-level quantities (like μ and σ)



Random Variables vs Observed Data

Three related (but different) ideas

- **Random variable:**

X = the random process before we observe anything

- **Observed data:**

x_1, x_2, \dots, x_n = the numbers we actually observe

- **Probability distribution:**

Describes how X behaves *before* data are collected

Key idea:

Data are *realizations* of a random variable, not the random variable itself.



What is a Random Variable?

Definition: Random Variable

A **random variable (r.v.)** assigns numerical values to the outcomes of a random phenomenon.

Notation:

- We use capital letters like X, Y, Z for random variables
- We use lowercase letters like x, y, z for specific values

Key idea: A random variable connects random outcomes to numbers and probabilities



Example: Rolling a Die

Suppose you roll a fair 6-sided die. Let X be the outcome of the roll.

Question 1: What is the probability distribution of X ?

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Note:

- Each outcome has equal probability (fair die)
- Probabilities sum to 1: $\sum_{x=1}^6 P(X = x) = 1$
- All outcomes are disjoint (mutually exclusive)



Probability Distributions: Quick Review

Rules for a Probability Distribution

A probability distribution must satisfy three rules:

1. The outcomes must be **disjoint** (mutually exclusive)
2. Each probability must be **between 0 and 1**
3. The probabilities must **sum to 1**

These are the same rules we learned in our probability unit!

Probability Distributions (Discrete)

For a **discrete** random variable, the **probability mass function (pmf)** is

$$f(x) = P(X = x)$$



Discrete vs. Continuous Random Variables

Discrete Random Variable

A **discrete r.v.** takes on:

- A finite number of values, OR
- A countably infinite number of values

Examples:

- Number of heads in 10 coin flips
- Number of students in a class
- Number of COVID cases per day

Continuous Random Variable

A **continuous r.v.** can take:

- Any real value in an interval
- Any value in a union of intervals

Examples:

- Height
- Blood pressure
- Time until an event occurs

Today's focus: Discrete random variables (continuous coming soon!)



Expected Value (Mean)

The **expected value** of a random variable is its long-run average value.

Definition: Expected Value of a Discrete R.V.

If X takes on outcomes x_1, \dots, x_k with probabilities $P(X = x_1), \dots, P(X = x_k)$, then:

$$E(X) = \mu = \sum_{i=1}^k x_i \cdot P(X = x_i)$$

The expected value is a **weighted average** where weights are probabilities.

Notation: $E(X)$ or μ (mu)



Example: Expected Value of Rolling a Die

Question 2: What is the expected outcome when rolling a fair die?

$$\begin{aligned} E(X) &= \sum_{x=1}^6 x \cdot P(X = x) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{1+2+3+4+5+6}{6} \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

Interpretation: If you rolled the die many times, the average would approach 3.5

Note: The expected value doesn't have to be a possible outcome!



Example: Unfair Die

Question 3: What if the die is not fair?

x	1	2	3	4	5	6
$P(X = x)$	$1/9$	$1/9$	$1/9$	$1/9$	$2/9$	$3/9$

$$\begin{aligned}E(X) &= 1\left(\frac{1}{9}\right) + 2\left(\frac{1}{9}\right) + 3\left(\frac{1}{9}\right) + 4\left(\frac{1}{9}\right) + 5\left(\frac{2}{9}\right) + 6\left(\frac{3}{9}\right) \\&= \left(\frac{1+2+3+4}{9}\right) + \left(\frac{10}{9}\right) + \left(\frac{18}{9}\right) \\&= \left(\frac{38}{9}\right) \\&= 4.222\dots\end{aligned}$$

The die is "loaded" toward higher values (mean is 4.22 vs. 3.5 for fair die)



Variance and Standard Deviation

Just like with data, we measure spread with variance and standard deviation.

Definition: Variance of a Discrete R.V.

If X has expected value $\mu = E(X)$, then:

$$\text{Var}(X) = \sigma^2 = \sum_{i=1}^k (x_i - \mu)^2 \cdot P(X = x_i)$$

The **standard deviation** is: $SD(X) = \sigma = \sqrt{\text{Var}(X)}$

Key idea: Squared deviations from the mean, weighted by probabilities



Example: Variance of a Fair Die

We found $E(X) = 3.5$. What is $\text{Var}(X)$?

$$\begin{aligned}\text{Var}(X) &= \sum_{x=1}^6 (x - 3.5)^2 \cdot P(X = x) \\&= (1 - 3.5)^2 \cdot \frac{1}{6} + (2 - 3.5)^2 \cdot \frac{1}{6} + \dots + (6 - 3.5)^2 \cdot \frac{1}{6} \\&= \frac{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25}{6} \\&= \frac{17.5}{6} \approx 2.92\end{aligned}$$

Standard deviation: $SD(X) = \sqrt{2.92} \approx 1.71$



Pause for a Quick Example

Apgar scores for newborns

Just after birth, each child is rated on the Apgar scale, with scores determined by color, heart rate, reflex irritability, muscle tone, and respiratory effort.

Let X = Apgar score. Historical data suggests the following probability distribution:

x	0	1	2	3	4	5	6	7	8	9	10
$P(X = x)$	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

Questions:

1. What is $E(X)$?
2. What is $\text{Var}(X)$?



Solution: Apgar Score Example (1/2)

x	0	1	2	3	4	5	6	7	8	9	10
$P(X = x)$	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

1. Expected value:

$$\begin{aligned}E(X) = \mu &= 0(0.002) + 1(0.001) + 2(0.002) + 3(0.005) + \dots + 8(0.25) + 9(0.12) + 10(0.01) \\&= 0 + 0.001 + 0.004 + 0.015 + \dots + 2.00 + 1.08 + 0.10 \\&= 7.17\end{aligned}$$

The average Apgar score is about 7.2.



Solution: Apgar Score Example (2/2)

x	0	1	2	3	4	5	6	7	8	9	10
$P(X = x)$	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

2. Variance:

$$\begin{aligned}\text{Var}(X) &= \sum_{x=0}^{10} (x - 7.17)^2 \cdot P(X = x) \\ &= (0 - 7.17)^2(0.002) + (1 - 7.17)^2(0.001) + \dots + (10 - 7.17)^2(0.01) \\ &= 0.103 + 0.038 + 0.053 + 0.087 + 0.201 + 0.188 \\ &\quad + 0.246 + 0.011 + 0.171 + 0.402 + 0.080 \\ &= 1.58\end{aligned}$$

Standard deviation: $SD(X) = \sqrt{1.58} \approx 1.26$

Interpretation: Most babies score between 6 and 9 (within 1 SD of the mean).



Linear Combinations of Random Variables

Often in research, we combine multiple measurements into a composite score.

Definition: Linear Combination

If X and Y are random variables and a and b are constants:

$$aX + bY$$

is a **linear combination** of the random variables.

Examples in biomedical research:

- Composite health scores (sum of multiple domains)
- Weighted risk scores
- Total costs (sum of different service types)



Expected Value of Linear Combinations

Theorem: Expected Value of Linear Combinations

If X and Y are random variables and a and b are constants:

$$E(aX + bY) = aE(X) + bE(Y)$$

and

$$E(aX + b) = aE(X) + b$$

Key properties:

- Expectation is **linear** - it distributes over addition
- Works whether or not X and Y are independent
- Constants factor out: $E(cX) = cE(X)$

This is intuitive: The average of a sum is the sum of the averages!



Example: Clinical Trial Health Score (Setup)

In a clinical trial, researchers measure three patient-reported outcomes (PROs):

- **Pain score (X_1)**: Scale 0-10, where 0 = no pain
 - $E(X_1) = 4, SD(X_1) = 2$
- **Mobility score (X_2)**: Scale 0-10, where 10 = full mobility
 - $E(X_2) = 6, SD(X_2) = 1.5$
- **Quality of life score (X_3)**: Scale 0-10, where 10 = excellent
 - $E(X_3) = 5, SD(X_3) = 2.5$

A **composite health score** is calculated as:

$$H = X_1 + X_2 + X_3$$

Assume the three scores are **independent** (one doesn't influence the others).



Example: Expected Composite Score

Question: What is the expected composite health score?

$$\begin{aligned}E(H) &= E(X_1 + X_2 + X_3) \\&= E(X_1) + E(X_2) + E(X_3) \\&= 4 + 6 + 5 \\&= 15\end{aligned}$$

Interpretation: On average, patients have a composite health score of 15 out of 30.



Variance of Linear Combinations

Theorem: Variance of Linear Combinations

If X and Y are **INDEPENDENT** random variables and a and b are constants:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

General formula when not independent

The general formula is

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

If independent, $\text{Cov}(X, Y) = 0$.

The standard deviation is the square root of the variance.



Example: Variance of Composite Score

Question: What is the variance and standard deviation of the composite health score?

Recall: $H = X_1 + X_2 + X_3$

- $\text{Var}(X_1) = 2^2 = 4$
- $\text{Var}(X_2) = 1.5^2 = 2.25$
- $\text{Var}(X_3) = 2.5^2 = 6.25$

$$\begin{aligned}\text{Var}(H) &= \text{Var}(X_1 + X_2 + X_3) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) \\ &= 4 + 2.25 + 6.25 \\ &= 12.5\end{aligned}$$

$$SD(H) = \sqrt{12.5} \approx 3.54$$



Example: Weighted Composite Score

Now suppose researchers want to weight the scores differently:

$$H_{\text{weighted}} = 0.5X_1 + 0.3X_2 + 0.2X_3$$

- Weights sum to 1, so this is a weighted average (0-10 scale), not a total score.
- Gives more weight to pain and mobility than quality of life.

Questions:

1. What is $E(H_{\text{weighted}})$?
2. What is $\text{Var}(H_{\text{weighted}})$?



Solution: Weighted Composite Score

1. Expected value:

$$\begin{aligned}E(H_{\text{weighted}}) &= E(0.5X_1 + 0.3X_2 + 0.2X_3) \\&= 0.5E(X_1) + 0.3E(X_2) + 0.2E(X_3) \\&= 0.5(4) + 0.3(6) + 0.2(5) \\&= 2 + 1.8 + 1.0 \\&= 4.8\end{aligned}$$

2. Variance:

$$\begin{aligned}\text{Var}(H_{\text{weighted}}) &= (0.5)^2\text{Var}(X_1) + (0.3)^2\text{Var}(X_2) + (0.2)^2\text{Var}(X_3) \\&= 0.25(4) + 0.09(2.25) + 0.04(6.25) \\&= 1.0 + 0.20 + 0.25 \\&= 1.45\end{aligned}$$

$$SD(H_{\text{weighted}}) = \sqrt{1.45} \approx 1.20$$



Key Takeaways: Linear Combinations

Expected value:

$$E(aX + bY) = aE(X) + bE(Y)$$

- Linear property
- No independence needed
- Constants factor out

Variance:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

- Requires independence
- Constants are squared
- Standard deviations don't simply add

Why this matters: Many biomedical measures are composite scores!



Part 2: Binomial Distribution

Section 3.2



Binary Outcomes

Many situations involve **two possible outcomes**:

- Flipping a coin: heads or tails
- Medical test: positive or negative
- Surgery outcome: success or failure
- Manufacturing: defective or non-defective

We call these **binary** or **Bernoulli** trials:

- One outcome is called a “success” (probability p)
- The other is called a “failure” (probability $q = 1 - p$)



Bernoulli Distribution

Definition: Bernoulli Random Variable

If X is a random variable that takes:

- Value 1 with probability p (success)
- Value 0 with probability $1 - p$ (failure)

Then X is a **Bernoulli random variable**.

Notation: $X \sim \text{Bernoulli}(p)$ or $X \sim \text{Bern}(p)$

Parameter: p is the probability of success (between 0 and 1)



Bernoulli: Mean and Variance

Theorem: Mean and Variance of Bernoulli R.V.

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

$$E(X) = p$$

$$\text{Var}(X) = p(1 - p)$$

Example with diabetes:

- Among US adults aged 65-74, about 20.3% have diabetes
- Define $X = 1$ if a randomly selected adult has diabetes, $X = 0$ otherwise
- $E(X) = 0.203$
- $\text{Var}(X) = 0.203(0.797) = 0.162$



From Bernoulli to Binomial

What if we repeat a Bernoulli trial multiple times?

Example: Randomly select 10 adults aged 65-74. How many have diabetes?

- Each person: Bernoulli(0.203)
- Total number with diabetes: Binomial(10, 0.203)

Key Relationship

$\text{Binomial}(n, p) = \text{Sum of } n \text{ independent Bernoulli}(p) \text{ trials}$



From Bernoulli to Binomial (continued)

- The **Bernoulli distribution** is a special case of the Binomial distribution where $n = 1$
 - Specifically:

$$\text{Binomial}(1, p) = \text{Bernoulli}(p)$$

- To get a **Binomial distribution**, we simply extend the scenario from a **single** trial to **multiple** independent trials.
 - If we conduct n independent Bernoulli trials with the same success probability p , the total number of successes across these n trials will follow a Binomial distribution



Binomial Distribution: Definition

Definition: Binomial Random Variable

X is a **Binomial random variable** if:

1. There are n independent trials
2. Each trial has two outcomes: success or failure
3. Probability of success is p (same for all trials)
4. X counts the total number of successes

Notation: $X \sim \text{Binomial}(n, p)$ or $X \sim \text{Binom}(n, p)$

Parameters: n (number of trials) and p (probability of success)

Possible values: $X \in \{0, 1, 2, \dots, n\}$



When NOT to Use a Binomial Model

Common Pitfalls

A Binomial model is **NOT appropriate** when:

- Trials are not independent
 - Example: sampling without replacement from a small population
 - (Probabilities change after each draw)
- Probability of success is not constant
 - Example: different subjects have different risk levels
 - (p varies across trials)

Rule of thumb: If independence or constant p is violated → not Binomial



Derivation by example

Approximately 20.3% of US adults aged 65-74 have diabetes.

Here, let's define

- "Success" (S) to be the event that a person has diabetes, then
- "Failure" (F) is the event that a person does not have diabetes.

Among US adults aged 65-74:

- $P(\text{has diabetes}) = 0.203 = p$
- $P(\text{no diabetes}) = 1 - p = 0.797$



Define an experiment

Suppose we conduct an experiment in which $n = 3$ US adults aged 65-74 are tested for diabetes.

- The testing of each adult represents a trial.
- The outcome from each trial is either a success (has diabetes) or a failure (does not have diabetes).
- The outcome (S/F) of any one trial is not influenced by earlier outcomes nor does it affect later outcomes (independent trials).
- The probability of having diabetes ($p = 0.203$) is the same for each adult sampled.



Possible outcomes

With 3 trials and 2 possible outcomes (S/F), there are 8 possible arrangements.

FFF	FSS
FFS	SFS
FSF	SSF
SFF	SSS



Define the rv

Let X = the number of adults that have diabetes among the $n = 3$ sampled.

Re-organize the possible arrangements:

- $X = 0 \iff FFF$
- $X = 1 \iff FFS \cup FSF \cup SFF$
- $X = 2 \iff FSS \cup SFS \cup SSF$
- $X = 3 \iff SSS$



Compute the probabilities

Case: $X = 0$ (no one has diabetes)

$$\begin{aligned} P(X = 0) &= P(FFF) \\ &\stackrel{\text{ind.}}{=} (1 - p)(1 - p)(1 - p) \\ &= (1 - p)^3 = 0.797^3 \approx 0.506 \end{aligned}$$

Because of the general multiplication rule for independent events, we simplified the event FFF into the product of $n = 3$ separate terms.



Compute the probabilities (continued)

Case: $X = 1$ (exactly 1 has diabetes)

- $FFS \implies (1 - p)(1 - p)p = p(1 - p)^2$
- $FSF \implies (1 - p)p(1 - p) = p(1 - p)^2$
- $SFF \implies p(1 - p)(1 - p) = p(1 - p)^2$

Substitute for p and $(1 - p)$ then for each $p(1 - p)^2 = (0.203)(0.797)^2 \approx 0.129$.

$$\begin{aligned} P(X = 1) &= P(FFS \cup FSF \cup SFF) \\ &= P(FFS) + P(FSF) + P(SFF) \\ &= 3p(1 - p)^2 \approx 3(0.129) = 0.387 \end{aligned}$$



Compute the probabilities (continued)

Case: $X = 2$ (exactly 2 have diabetes)

Similar reasoning gives

- $FSS \implies (1 - p)pp = p^2(1 - p)$
- $SFS \implies p(1 - p)p = p^2(1 - p)$
- $SSF \implies pp(1 - p) = p^2(1 - p)$

Each of the three parts is approximately 0.033

$$\begin{aligned} P(X = 2) &= P(FSS \cup SFS \cup SSF) \\ &= P(FSS) + P(SFS) + P(SSF) \\ &= 3p^2(1 - p) \approx 3(0.033) = 0.099 \end{aligned}$$



Compute the probabilities (continued)

Case: $X = 3$ (all 3 have diabetes)

$$\begin{aligned}P(X = 3) &= P(SSS) \\&= ppp \\&= p^3 = (0.203)^3 \approx 0.008\end{aligned}$$



The Pattern Emerges

Looking at our calculations:

X	Number of arrangements	Probability
0	1	$1 \times (0.203)^0(0.797)^3 = 0.506$
1	3	$3 \times (0.203)^1(0.797)^2 = 0.387$
2	3	$3 \times (0.203)^2(0.797)^1 = 0.099$
3	1	$1 \times (0.203)^3(0.797)^0 = 0.008$

Pattern:

$$P(X = k) = (\text{number of arrangements}) \times p^k(1 - p)^{n-k}$$

Question: How do we count the number of arrangements for any n and k ?



Counting Arrangements: "n choose k"

The number of ways to arrange k successes among n trials is:

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

This is called the **binomial coefficient** or "n choose k"

Reminder: $n!$ (read "n factorial") means $n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1$

- Example: $3! = 3 \times 2 \times 1 = 6$
- Special case: $0! = 1$ (by definition)

For our example with $n = 3$:

- $\binom{3}{0} = \frac{3!}{0!3!} = \frac{6}{1 \times 6} = 1 \checkmark$
- $\binom{3}{1} = \frac{3!}{1!2!} = \frac{6}{1 \times 2} = 3 \checkmark$
- $\binom{3}{2} = \frac{3!}{2!1!} = \frac{6}{2 \times 1} = 3 \checkmark$
- $\binom{3}{3} = \frac{3!}{3!0!} = \frac{6}{6 \times 1} = 1 \checkmark$



Binomial distribution

The final rule is

Distribution of a Binomial random variable

Let X be the total number of successes in n independent trials, each with probability p of a success. Then probability of observing exactly k successes in n independent trials is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0, 1, 2, \dots, n$$

Components:

- $\binom{n}{k}$: number of ways to arrange k successes in n trials
- p^k : probability of k successes
- $(1 - p)^{n-k}$: probability of $n - k$ failures



Verify with our example

Does our formula work for $n = 3, p = 0.203$?

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

k	Formula	Probability
0	$\binom{3}{0} (0.203)^0 (0.797)^3$	0.506
1	$\binom{3}{1} (0.203)^1 (0.797)^2$	0.387
2	$\binom{3}{2} (0.203)^2 (0.797)^1$	0.099
3	$\binom{3}{3} (0.203)^3 (0.797)^0$	0.008

Matches what we calculated by hand! ✓



Now extend to 10 adults

Now we can easily calculate probabilities for $n = 10$ adults:

Setup: $X \sim \text{Binomial}(n = 10, p = 0.203)$

Without the formula: We'd need to enumerate $2^{10} = 1024$ possible arrangements!

With the formula: We can calculate any probability directly:

$$P(X = 4) = \binom{10}{4} (0.203)^4 (0.797)^6 = \frac{10!}{4!6!} (0.203)^4 (0.797)^6 \approx 0.091$$

The binomial formula saves us from tedious counting!



R Makes This Even Easier!

We can also use R to calculate binomial probabilities:

```
1 # Verify our n=3 calculations  
2 dbinom(x = 0:3, size = 3, prob = 0.203)  
[1] 0.506261573 0.386842281 0.098530719 0.008365427
```

```
1 # Calculate P(X = 4) for n=10  
2 dbinom(x = 4, size = 10, prob = 0.203)  
[1] 0.09140151
```

This is especially helpful when you need many probabilities or when n is large!



Binomial: Mean and Variance

Theorem: Mean and Variance of Binomial R.V.

If $X \sim \text{Binomial}(n, p)$, then:

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p)$$

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

Intuition:

- If each trial has probability p of success, and we do n trials, we expect np successes on average
- The variance formula comes from summing n independent Bernoulli variances



Example: Diabetes in Older Adults (10 people)

Approximately 20.3% of US adults aged 65-74 have diabetes. Suppose we randomly select 10 adults in this age group (independently).

Let X = number with diabetes among the 10 people

Setup: $X \sim \text{Binomial}(n = 10, p = 0.203)$

Question 1: What is the expected value of X ?

$$E(X) = np = 10(0.203) = 2.03$$

Question 2: What is the standard deviation of X ?

$$SD(X) = \sqrt{np(1 - p)} = \sqrt{10(0.203)(0.797)} = \sqrt{1.618} \approx 1.27$$



Example: Exactly 4 with Diabetes

Question 3: What is the probability that exactly 4 of the 10 have diabetes?

$$\begin{aligned} P(X = 4) &= \binom{10}{4} (0.203)^4 (0.797)^6 \\ &= \frac{10!}{4! \cdot 6!} (0.203)^4 (0.797)^6 = 210 \times 0.0017 \times 0.256 \approx 0.091 \end{aligned}$$

In R:

```
1 dbinom(x = 4, size = 10, prob = 0.203)
[1] 0.09140151
```

- `dbinom` gives the **pmf**: $P(X = x)$ for a discrete distribution. (probability at a specific value)
- R uses “d” for these functions to stand for “density”; for discrete distributions it returns the probability mass.



Example: At Most 3 with Diabetes

Question 4: What is the probability that at most 3 have diabetes?

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

We could calculate each term, but there's an easier way!

In R:

```
1 pbinom(q = 3, size = 10, prob = 0.203)
[1] 0.8737821
```

The `p` in `pbinom` stands for "cumulative probability", $P(X \leq k)$

Result: About 87% chance that 3 or fewer have diabetes



Example: At Least 5 with Diabetes

Question 5: What is the probability that at least 5 have diabetes?

$$P(X \geq 5) = P(X = 5) + P(X = 6) + \cdots + P(X = 10)$$

Approach 1: Use complement rule

$$P(X \geq 5) = 1 - P(X \leq 4)$$

```
1 1 - pbinary(q = 4, size = 10, prob = 0.203)
[1] 0.03481642
```

Approach 2: Use `lower.tail = FALSE`

```
1 pbinary(q = 4, size = 10, prob = 0.203, lower.tail = FALSE)
[1] 0.03481642
```

Result: About 3.4% chance that 5 or more have diabetes



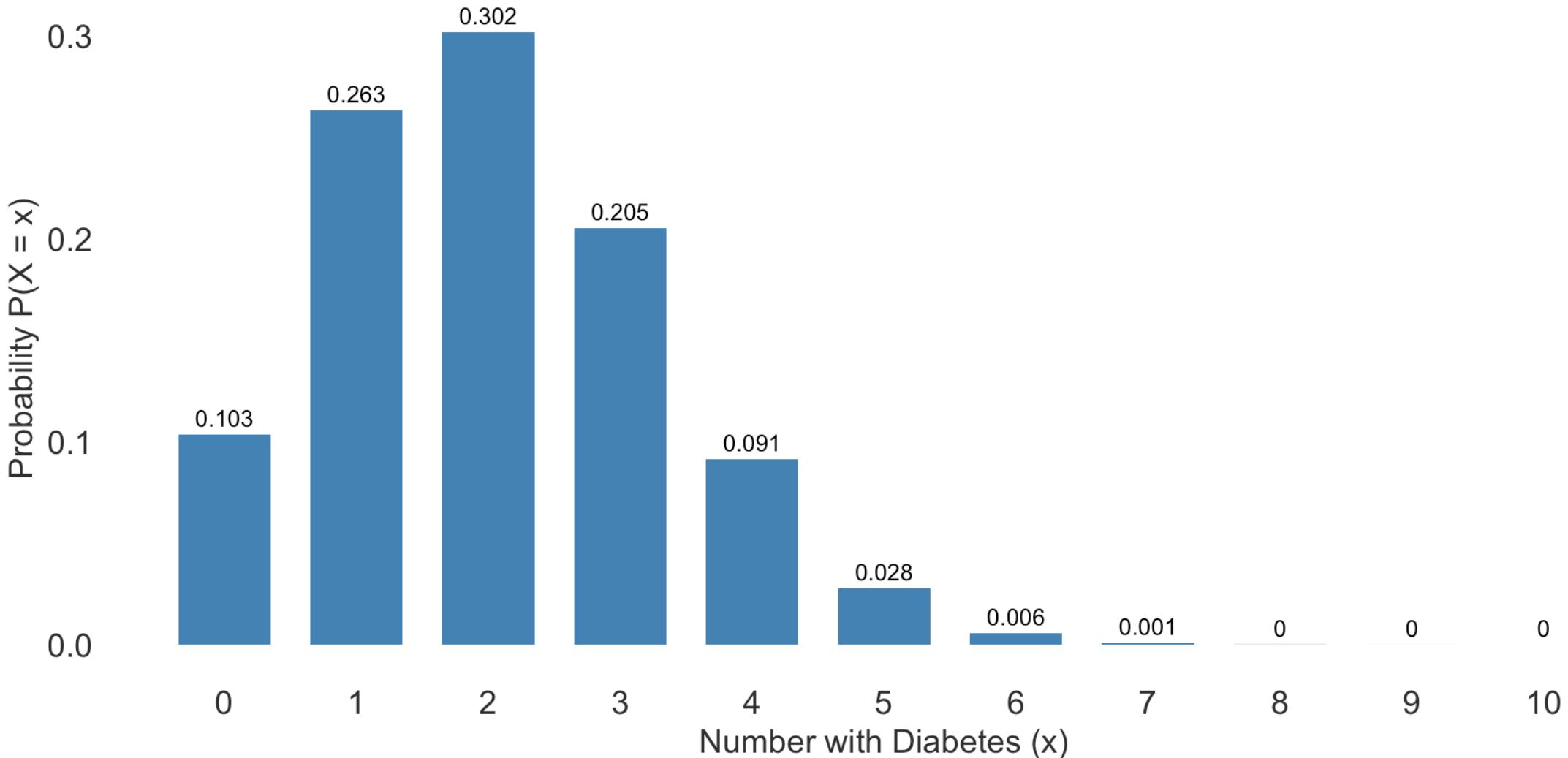
R Functions for Binomial Distribution

Function	What it does	Example
<code>dbinom()</code>	Probability at specific value: $P(X = k)$	<code>dbinom(x = 4, size = 10, prob = 0.203)</code>
<code>pbinom()</code>	Cumulative probability: $P(X \leq k)$	<code>pbinom(q = 3, size = 10, prob = 0.203)</code>
<code>qbinom()</code>	Quantile: Find k for given probability	<code>qbinom(p = 0.5, size = 10, prob = 0.203)</code>
<code>rbinom()</code>	Generate random samples	<code>rbinom(n = 100, size = 10, prob = 0.203)</code>



Visualizing the Binomial Distribution

Binomial Distribution: $n = 10$, $p = 0.203$



Key Takeaways

Random Variables:

- Connect random outcomes to numerical values and probabilities
- Expected value: weighted average using probabilities
- Variance: measures spread around the expected value
- Linear combinations: $E(aX + bY) = aE(X) + bE(Y)$, but $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ (if independent)

Binomial Distribution:

- Models number of successes in n independent trials
- Parameters: n (trials) and p (success probability)
- $E(X) = np$, $\text{Var}(X) = np(1 - p)$
- Use R functions: `dbinom()`, `pbinom()`, etc.



Looking Ahead

Wednesday:

- Normal distribution
- Central Limit Theorem
- Sampling distributions

Next steps:

- Review textbook sections 3.1-3.2
- Work on HW 2 (due Tuesday, January 27)
- Practice calculating binomial probabilities in R



Questions?

