PHIL 7001: Fundamentals of AI, Data, and Algorithms

Week 4
Distribution Theory

Boris Babic,
HKU 100 Associate Professor of Data Science, Law and Philosophy

Today

Learning goals

- Introduction to Probability Distributions
- Discrete vs. Continuous Distributions
- Binomial Distribution and Examples
- Normal Distribution and its Role
- Properties of the Normal Distribution

Last week

Review of last week

- Introduction to probability
- Sample space, outcomes, events
- Properties of sets (union, intersection, complement)
- Laws of probability
- Bayes' Rule

Random Variables

Recap on sample space

- In working with probability questions, it is important to know our sample space!
- What are all the possible outcomes? Ex: a die can land on 1...6. What are the events of interest? Ex: the likelihood that a die lands on an even number. Even number $= \{2, 4, 6\}$.

Calculating Probability by Counting Outcomes from the Sample Space Ω

- We can use this method if:
  - All outcomes of the sample space are equally likely.
  - The sample space Ω is finite.
- Let A be an event defined on the sample space Ω.
- The probability of event A can be calculated as:

$$P(A) = \frac{\text{number of outcomes satisfying event } A}{\text{Total number of outcomes in } \Omega}$$

Example 1: What is the probability of getting a head if a fair coin is tossed?

- The sample space $\Omega = \{H, T\}$
- Event: We are looking for a head.
- Out of the two outcomes, one satisfies our condition.
- Hence, $P(H) = \frac{1}{2}$.

Random variables

- We will use capital end of alphabet letters, $...W, X, Y, Z$, to indicate random variables.
- Random variables map outcomes in the sample space to natural or real numbers.
- In this sense, they are a function: $X : S \to \mathbf{R}$.
- For example: If we are modeling a coin toss, we might say we have a random variable X, which can take the value 1 for heads, or 0 for tails.
- If we are interested in students' heights, then we can assume $X \in \mathbf{R}^+$.
- Random variables can be discrete, or continuous.
- We will use capital X to denote the random variable and lowercase x to indicate values it might take.

Probability Distributions

Probability distributions

- A probability distribution is a function f of the random variable X: $f(X)$
- It maps values of the random variable to probabilities.
- This function can only take values between 0 and 1.
- Also, this function is additive: for two independent events, the probability of their sum is the sum of their probabilities.

  Ex: Pr (die lands on 4) $+$ Pr (die lands on 6) $=$ Pr (die lands on 4 or die lands on 6).
- Finally, the total probability of the whole space must sum to 1.
- **Example:** Imagine you roll a fair six-sided dice. The outcomes are the numbers 1 through 6. Because the dice is fair, each number has an equal probability of $\frac{1}{6}$. The probability distribution can be represented as:

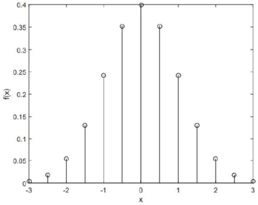
Outcome	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Two Main Types of Distributions

1. **Discrete Probability Distribution:** For discrete random variables. Outcomes take on distinct/separate values.
2. **Continuous Probability Distribution:** For continuous random variables. Outcomes take on a range of values without gaps.

Discrete Probability Distribution

Definition: In the discrete case, it tells us how probable it is that the random variable X will take a specific value x. We often denote the function f with $\Pr$. Possible values are countable and distinct.
Example: Tossing a coin, Rolling a die.
Visualization: Probability mass function (PMF).

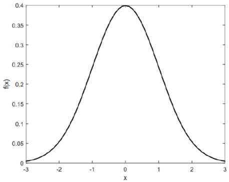


Characteristics:

- Probability for any specific value is non-negative: $P(X = x) \geq 0$.
- Sum of all probabilities is 1: $\sum P(X = x) = 1$.
- Example: Rolling a fair six-sided die. Each number 1-6 has a $\frac{1}{6}$ chance of occurring. All possibilities summed together equal 1.

Continuous Probability Distribution

Definition: In the continuous case, it tells us how probable it is that the random variable X will fall within some specified region between a and b.
Visualization: Probability density function (PDF).



Characteristics:

- Probability that X takes on any specific value is always 0.
- Area under the entire curve (PDF) is 1.
- To find the probability that X lies in a range of values, calculate the area under the curve for that range.
- Example: Height of adult men being between 5'8" and 6'2".

Binomial Distribution

Binomial Distribution

INTERESTED IN X

X = # of "Heads" in *n* tosses of a fair coin

X = # of times a Thumbtack will land on its flat head in *n* tosses

X = # of consumers in a sample of *n* people who prefer to use "Brand A" of a product versus all other brands

X = # of defective items in a lot of *n* items

X = # of job offers after interviews with *n* companies

Binomial Distribution

BINOMIAL DISTRIBUTION IF

1. ***n*** **trials such that each trial has only two possible outcomes: "Success" or "Failure" (*Bernoulli* trials)**
2. **P(Success) = *p* is the same for all trials**

   P(Failure) = 1 - *p* = *q*
3. **All trials are independent.**

Interested in the probability of observing

X = # of Successes in *n* trials

Binomial Distribution

- Suppose we are tossing a coin once, and we want to know the probability that it lands on heads (p), or tails ($1-p$).
- In general, we can write this as,

$$p^x(1-p)^{1-x}$$

- If we toss the coin n times, then this becomes

$$p^x(1-p)^{n-x}$$

- But we must account for the number of different ways we can observe x successes in n experiments.
- Example: What is the probability of observing 2 heads in 3 tosses? First, we have $p^2(1-p)^{3-2}$. If the coin is fair, then this is $0.5^2(1-0.5)^1 = 0.5^3 = 0.125$.
- But there are three ways to observe two heads in three tosses of a coin: HHT, HTH, THH. So we need to multiply 0.125 by $\binom{3}{2} = 3$. This is 0.375.
- So the full binomial distribution is given by,

$$\mathrm{P}(X = x) = \binom{n}{x} p^x(1-p)^{n-x},$$

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

Exercise

Exercise: In a factory, 95% of products pass quality control. If you randomly select 20 products, what's the probability that exactly 18 of them pass the quality control?

Exercise

Exercise: In a factory, 95% of products pass quality control. If you randomly select 20 products, what's the probability that exactly 18 of them pass the quality control?

Answer:

$$\mathrm{P}(X = 18) = \binom{20}{18}(0.95)^{18}(0.05)^{2} \approx 0.2029$$

Exercise

Exercise: You're taking a multiple-choice test with 10 questions. Each question has 4 choices and only one is correct. If you guess on each question, what's the probability of getting at least 8 questions correct?

Take 5 minutes to work on this in groups of 2.

Exercise

Exercise: You're taking a multiple-choice test with 10 questions. Each question has 4 choices and only one is correct. If you guess on each question, what's the probability of getting at least 8 questions correct?

Take 5 minutes to work on this in groups of 2.

Answer:

$$\mathrm{P}(X \geq 8) = \mathrm{P}(X = 8) + \mathrm{P}(X = 9) + \mathrm{P}(X = 10)$$

$$= \binom{10}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2 + \binom{10}{9}\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)^1 + \binom{10}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^0$$

$$\approx 0.00048$$

Exercise Solution using R

- In R, the function **dbinom(x,n, p)** returns the probability of x successes in n trials, given probability p.
- So you can use the code:

```
dbinom(x=8, size=10, prob=.25) +
dbinom(x=9, size=10, prob=.25) +
dbinom(x=10, size=10, prob=.25)
```

- More efficiently, the function **pbinom(x, n, p, lower.tail=FALSE)** returns the probability to the right of x, given n and p. Hint: If you do not include lower.tail=FALSE then it will return the probability to the left of x, but including x.
- Be careful: lower.tail=FALSE returns $Pr(X > x)$. But lower.tail=TRUE returns $Pr(X \leq x)$. So just remember this, depending on whether you want to include x or not.
- So you can simply write:

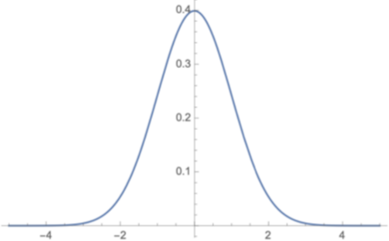
```
pbinom(7, 10, 0.25, lower.tail=FALSE)
```

- Notice that her we wrote 7, and not 8. Since the question said "at least 8" we want to include 8 in our calculation.

Normal Distribution

Normal Distribution

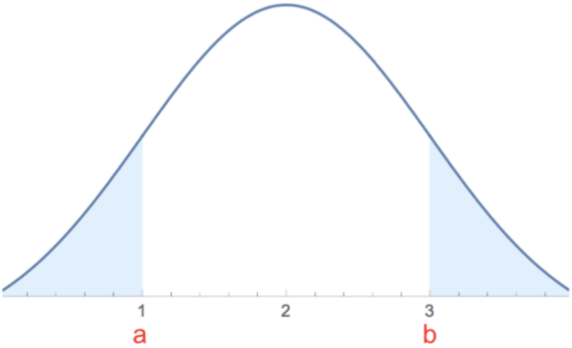
- Most important probability distribution you will encounter (due, in part, to the central limit theorem).
- This distribution belongs to the exponential family of distributions, and it has two parameters, its average μ and standard deviation σ.
- Represented by the famous "bell curve": symmetric around its mean



- Given by

$$f(x|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

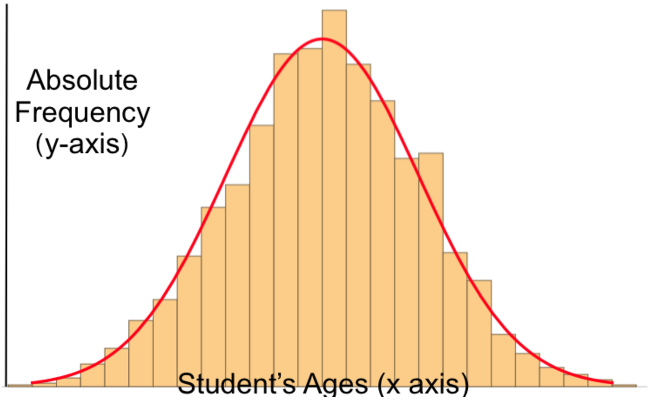
Normal Distribution



The probability that our random variable is between a and b is given by the area under the curve between those two points:

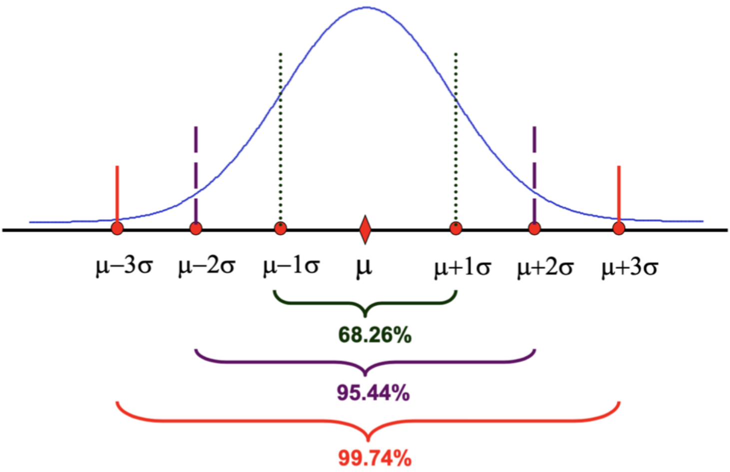
$$\Pr(a < x < b) = (2\pi\sigma^2)^{-1/2} \int_a^b e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

Normal Distribution

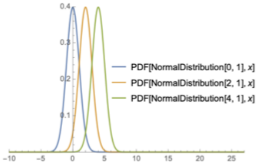


- Variables that have a normal distribution are ubiquitous in real life, provided we have enough data.
- Age of HKU students, height of HKU students.

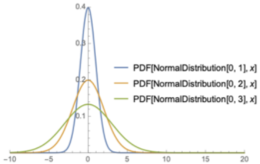
Empirical Rule



Personality of Normal Parameters



- As the mean changes, the location of the bell shifts
  To the left (for smaller means)
  To the right (for larger means)



- As the standard deviation changes, the bell becomes
  taller and thinner (for smaller standard deviations)
  shorter and thicker (for larger standard deviations)

The pnorm Function in R

- Suppose that the test scores of a course exam at HKU are normally distributed with a mean of 72 and a standard deviation of 15.2. What is the probability that a randomly chosen student received above 84?
- `pnorm(84, mean=72, sd=15.2, lower.tail=FALSE)`.
- Approximately 21%.
- We use `lower.tail=FALSE` in order to get the area from x to ∞.
- If you want the area to the left x, then as before, do not include lower.tail=FALSE.
- Example: Find the percentage of otters that weigh less than 33 kilograms in a population with mean = 40 and sd = 8.
- pnorm(33, mean=40, sd = 8) = 0.19.

The pnorm Function in R

- The weekly salaries of the employees of a large corporation are assumed to be normally distributed with mean \$450 and standard deviation \$40.
- What is the probability that a randomly chosen employee earns more than \$500 per week?

The pnorm Function in R

- The weekly salaries of the employees of a large corporation are assumed to be normally distributed with mean \$450 and standard deviation \$40.
- What is the probability that a randomly chosen employee earns more than \$500 per week?

  - pnorm(500, mean=450, sd=40, lower.tail=FALSE)
  - approximately 10%.

Tips

- Probabilities correspond to areas
- Probabilities sum to 1: $Pr(X < k) = 1 - Pr(X > k)$
- Symmetry: $Pr(X < -k) = Pr(X > k)$
- For intervals, use subtraction:
  $Pr(a < X < b) = Pr(X < b) - Pr(X < a)$

Summary

- Binomial: Discrete, characterized by two parameters n, p.
- Normal: Continuous, characterized by two parameters, μ, σ.
- Binomial in R: pbinom(x, n, p), dbinom(x, n, p, lower.tail=FALSE).
- Normal in R: dbinom(x, mu, sigma, lower.tail=FALSE).
- Now you know how to calculate normal and binomial probabilities for any event without ever having to use one of those complicated look up tables at the back of statistics textbooks! WOW!!

Next class

- Inference: estimates, hypothesis tests, p-values
- Bayesian inference

Resources

- Whenever you need help with R, I highly recommend to Google search the package documentation. For example, for the binom function in R, everything can easily be found here: `https://www.rdocumentation.org/packages/stats/versions/3.3/topics/Binomial`