Intuition for Stats

🔍 CLT Red Flags in Word Problems

1. Sampling from a population repeatedly

Look for phrases like:

- "A sample of size n is taken..."
- "The average of 50 measurements..."
- "Suppose 100 items are selected at random..."

These suggest you're working with **sample statistics**, which is where the CLT comes in.

2. You're asked about a probability involving a sample mean

Key phrases:

- "What is the probability that the *average* weight is more than 150 lbs?"
- "Find the probability that the *mean* delivery time is less than 3 days."
- "Determine the probability that the *sample mean* falls between..."

This is a clear sign you need to apply CLT to the **distribution of sample means**.

3. The population distribution is NOT normal or unknown

CLT becomes relevant **especially** when the population is not normally distributed (or you aren't told). The CLT tells you the **sampling distribution of the sample mean** will be approximately normal if the sample size is large.

4. Sample size n is "large" (typically $n \geq 30$)

Even if the population isn't normal, CLT allows you to assume normality for the sample mean when n is large. If it's mentioned that:

- "The sample size is 40."
- "100 observations were taken."

...CLT is likely in play.

5. You’re given population mean and standard deviation

Phrases like:

- "The population has mean $\mu=100$and standard deviation $\sigma=15$”
- "The population standard deviation is known..."

Mean + standard deviation + sample size is a classic setup for applying:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Which is the **CLT Z-score formula for sample means**.

🚫 CLT Usually Not Involved When:

- You're dealing with **a single observation** from the population.
- You're finding probabilities about **counts**, not **averages**.
- You're working with exact small samples from small populations (may require **exact binomial**, **hypergeometric**, or **t-distribution**).

✅ TL;DR: Use CLT when...

Clue	Meaning
"Sample mean" or "average"	CLT is likely
Large sample size $n \geq 30$	Approx. normal by CLT
Population distribution not normal or not mentioned	CLT helps you assume normality
You're given μ, σ, and n, and asked about probability	Classic CLT setup

---

🔍 How to Recognize When to Use Conditional Probability (Plain)

Conditional Probability Formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

✅ Clues that *plain conditional probability* is needed:

1. **Problem talks about "given that..."**
   - Look for wording like:

   > "What is the probability that the person is a smoker given that they are over 50 years old?"

2. **You know/are given P(A and B) and P(B)**
   - The problem provides both probabilities directly, and you just need to use the formula above.
3. **No "updating" beliefs after new information**
   - You're not "reversing" known probabilities (like Bayes); you're directly finding how two events relate.

🚫 Not conditional probability if:

- You're trying to **update** probabilities based on new evidence (that's Bayes').
- No event is "given" in the wording.

🔍 How to Recognize When to Use Bayes' Theorem

Bayes' Theorem Formula:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

✅ Clues that Bayes' Theorem is needed:

1. **You're given "reverse" information**
   - The problem tells you about P(B|A) (e.g., test positive if sick), but asks for P(A|B) (e.g., sick if tested positive).
   - Common phrasing:

> "Given that the test was positive, what is the probability that the person has the disease?"

2. **You have prior ("before") probabilities and you must "update" after new evidence**
    - For example:
        - Prior: 2% of people have a disease.
        - Evidence: Positive test.
        - Question: Now, what's the probability they *actually* have the disease?
3. **Problem mentions "false positives", "true positives", or "testing"**
    - Testing problems (medical tests, lie detectors, defect tests) very often use Bayes.
4. **Multiple possible causes or sources**
    - Problem says:

    > "There are 3 machines, each produces different % of defects. Given that an item is defective, what is the probability it came from machine 1?"

    - → **Bayes' Theorem needed**.

🔥 Simple Comparison

If you are asked for...	Use
Probability of A given B, and you know joint and marginal probabilities	Conditional Probability
Update your belief about A after observing B, especially in diagnosis, testing, or backward reasoning	Bayes' Theorem

---

✅ Quick Visual Memory Trick

- **Direct "given that"** → **Conditional probability** (simple formula).
- **Reverse thinking or updating beliefs** → **Bayes' theorem** (more involved).

Probability questions involving **marbles or balls** typically fall under **classical probability** where outcomes are **finite, countable, and equally likely**. Here's a breakdown of the logic you can follow:

🔁 General Steps for Marble/Ball Probability Problems:

1. **Understand the Setup**:
    - How many marbles/balls total?
    - What are the colors/types?
    - Are the draws with or without replacement?
2. **Define the Events**:
    - What event's probability are you trying to compute? (e.g., both red, at least one blue, etc.)
3. **Count Favorable Outcomes**:
    - How many ways can the desired event occur?
4. **Count Total Outcomes**:
    - Total number of possible outcomes (depends on replacement and number of draws).
5. **Apply the Basic Formula**:

$$\text{Probability} = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

📌 Key Variations and What to Look For:

Type of Problem	Clue Words	What to Do
Without Replacement	“not replaced”, “removed”	The total and probabilities change after each draw.
With Replacement	“replaced”, “put back”	Probabilities stay the same for each draw.
Combinations	“choose 2”, “randomly select k”	Use $\binom{n}{k}$ for selecting without order.
Conditional Probability	“given that”, “if known that”	Use conditional probability formula: $P(A \mid B)=P(A \cap B)/P(B)$

Expected Value (advanced)	“on average”, “expected number”	Multiply each outcome by its probability and sum.

🧠 Example Logic (Simple):

Q: A bag has 5 red and 3 blue balls. Two balls are drawn *without replacement*. What’s the probability both are red?

- Total balls = 8
- Drawing 2 balls without replacement:

$$P(\text{both red}) = \frac{5}{8} \cdot \frac{4}{7} = \frac{20}{56} = \frac{5}{14}$$

1. 🎯 Bernoulli Random Variable

Definition:

Single trial with only **two outcomes**: success (1) or failure (0).

Recognize when:

- You are doing **just one trial**.
- Only two outcomes ("success/failure", "yes/no", "win/lose").

Key words:

- "One attempt"
- "Did the customer buy the product: yes or no?"
- "Did the patient survive: yes or no?"

Important:

- If it's only **one** trial → **Bernoulli**.

2. 🎯 Binomial Distribution

Definition:

Repeated **n** independent **Bernoulli trials**, counting the number of successes.

Recognize when:

- There are **multiple independent trials** (fixed number n).
- Each trial has **only two outcomes**.
- **Same probability** of success each time.

Key words:

- "Fixed number of trials"
- "5 customers come in. What is the probability 3 buy something?"
- "In 10 shots, probability of scoring exactly 7 goals."

Important:

- If there's **n independent, identical trials** → **Binomial**.
- Formula involves combinations:

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

---

3. 🎯 Poisson Distribution

Definition:

Counts the number of events happening in a **fixed interval** (time, area, volume), where events occur **independently** at a constant average rate.

Recognize when:

- **No fixed number of trials** — events just happen randomly.
- Describing events **over time or space**.
- Events are rare.

Key words:

- "per hour", "per day", "per km", "per page"
- "On average, 3 accidents occur per week."
- "How many calls are received at a call center in an hour?"

Important:

- **Poisson** is for **event counts** in **continuous domain** (time, space).

- Formula:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where λ= average number of events.

---

4. 🎯 Uniform Distribution

(Discrete Uniform)

Definition:

Each of a **finite number** of outcomes is **equally likely**.

Recognize when:

- All outcomes have **equal probability**.
- Picking at random from a set.

Key words:

- "Roll a die" (faces 1–6 are equally likely)
- "Choose a card at random"

(Continuous Uniform)

Definition:

All real numbers between two values are **equally likely**.

Recognize when:

- Picking a **real number** at random between two limits.
- Continuous outcomes (not just countable).

Key words:

- "Arrival time is uniformly distributed between 2 PM and 4 PM."
- "Temperature is uniformly distributed between 20°C and 30°C."

Important:

- For continuous:

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

---

5. 🎯 Normal Distribution (Gaussian)

Definition:

A bell-shaped, symmetric distribution used for many natural phenomena.

Center = mean, **spread** = standard deviation.

Recognize when:

- The data is **continuous**.
- The problem talks about **averages**, **measurements**, **heights**, **weights**, **test scores**, etc.
- Symmetric around a center value.

Key words:

- "Normally distributed"
- "The height of students is normally distributed with mean 170 cm and standard deviation 10 cm."
- "Assume normal distribution."

Important:

- Z-scores are often used:

$$Z = \frac{X - \mu}{\sigma}$$

🎯 Bonus Tip

If a problem says "mean = μ, variance = σ^2" and **large number of trials or samples** → **CLT (Central Limit Theorem)** is likely involved, even if the original distribution is not normal!