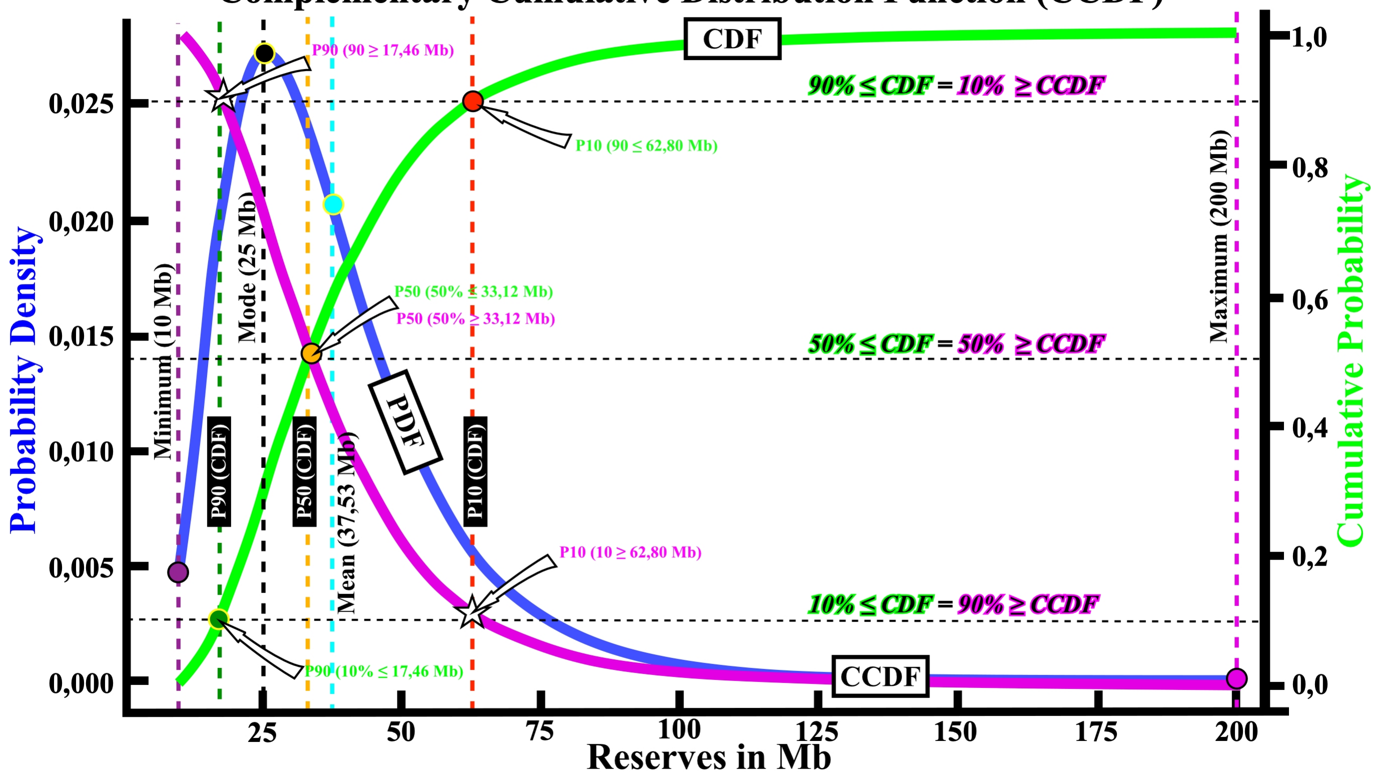
Understanding Probabilistic Reserves

A Manager's Guide to Key Metrics and Insights

Probability Density Function (PDF)
Cumulative Distribution Function (CDF)
Complementary Cumulative Distribution Function (CCDF)



The green curve (CDF) inherently describes the probability that reserves are less than or equal to a given value. The complementary probability is 1 - CDF, which is the Complementary Cumulative Distribution Function (CCDF), shown in purple. The CCDF describes the probability that the reserves are greater than or equal to a given value.

Ccramez, 2003, 2024

1- Introduction

Probabilistic reserves estimation is a statistical approach to evaluate and understand petroleum (oil or gas) reserves by taking into account uncertainties in geological, engineering and economic factors. This method generates a probability distribution for reserves, which allow us to characterise reserves using terms such as P90, P50, P10 (Probabiliies) and statistical descriptors such as Maximum, Mode, Mean and Minimum.

To illustrate how these terms are calculated and interpreted, these notes provide essential definitions used in most oil companies, relationships and working examples.

2- Key Definitions

2.1- Probability Density Function (PDF)

The PDF gives the density of probability at a given amount of hydrocarbons, but not the exact probability. For continuous variables, the probability of a single point is zero. Instead, the PDF is used to calculate probabilities over a range of values.

Mathematically, the probability of a given amount of hydocarbons falling within an interval [a,b] is the area under the PDF curve between a and b:

$$P\ (a \leq X \leq b) = \int_{(a-b)} f(x)\ dx$$

where f(x) is the PDF.

- **The PDF curve indicates how concentrated or disersed the probabilities are.**
- **The peak of the PDF corresponds to the most likely value (Mode), but it does not represent absolute probabilities.**
- **If the PDF for hydrocarbon reserves shows a peak around 125 Mb, this indicates that 125 Mb is the most likely reserve value.**
- **The PDF does not tell us directly the probability that, for example, the reserves are exactly 25 Mb (which is zero for continuous data). Instead, it tells us the relative likelihood compared to other values.**

2.2- Cumulative Distribution Function (CDF):

The CDF represents the probability that rthe amount of hydrocarbons will be less than or equal to a given value. For example, a P90 value indicates that there is a 10% probability that the amount of hydrocarbons will be equal to or less than a given value.

- **The CDF is calculated by integrating the PDF:**

$$F(x) = \int_{(x\ -\infty)} f(t)\ dt$$

- **Conversely, the PDF is the derivative of the CDF:**

$$f(x) = (d/(dx)\ F(x)$$

- **The total area under the PDF curve is equal to 1, representing the entire probability space.**

2.3- Complementary Cumulative Distribution Function (CCDF)

The CCDF, also known as the "Survival Function", represents the probability that the variable (amount of HC) is greater than or equal to a given value. Mathematically:

$$CCDF(x) = 1\text{- } CDF(x)$$

where CDF(x) is the Cumulative Distribution Function.

It is called «complementary» because it complements the CDF, as the two together account for all probabilities:

$$CDF(x) + CCDF(x) = 1$$

The CCDF is widely used in reserve evaluation because it indicates the probability of exceeding reserve thresholds, which is critical in high-risk decision-making processes. For example, a P90 value (see: 1. Definition of Percentiles in the Appendix) indicates in the CCDF context that there is a 90% probability that reserves will be greater than or equal to a certain value.

2.4- Probabilistic mesures: P90, P50, and P10

Probabilistic measures are derived from a statistical distribution (e.g. log-normal distribution) that quantify the amount of hydocarbons based on confidence levels. These measures can be determined using the Cumulative Distribution Function (CDF) or the Complementary Cumulative Distribution Function (CCDF). (see: 1. Definition of Percentiles in the Appendix).

2.4.1- Cumulative Distribution Function (CDF):

The CDF represents the probability that the volume of hydrocarbons is less than or equal to a specific value. On this basis, the estimated amount of hydrocarbons are defined as follows:

- **P90 :**
  The volume of hydrcarbons with a 10% probability of being less than or equal to a given value. This represents a conservative estimate of reser-ves, with a high degree of certainty.

☑ **P50:**

The volume of hydrocarbons with a 50% probability of being less than or equal to a given value. This is considered the best estimate or Median, and represents an equal chance of reserves being above or below this value.

☑ **P10:**

The volume of hydrocarvons with a 90% probability of being less than or equal to a given value. This represents an optimistic estimate with a lower degree of certainty.

2.4.2- Complementary Cumulative Distribution Function (CCDF):

The CCDF represents the probability that the volume of hydrcarbons is greater than a given value. Based on the CCDF, reserve estimates are defined as follows:

☑ **P90:**

The volume of hydrocarbons with a 90% probability of being greater than or equal to a given value. This represents a conservative estimate, with a high degree of certainty that volume is below this value.

☑ **P50:**

The volume of hydrocarbons with a 50% probability of being greater than or equal to a given value. This corresponds to the "best estimate » or Median and indicates an equal chance of tha volume being above or below this value.

☑ **P10:**

The volume of hydrocarbons with a 10% probability of being greater than or equal to a given value. This is an optimistic estimate, with a lower level of certainty, representing the potential for higher volume.

2.4.3- Clarifications and Key Points:

(i) As you may have noticed, I am deliberately avoiding the terms "resources"and "reserves" to prevent potential misunderstandings. The term "volume" or 'amount' of hydrocarbons is intentionally less restrictive, as it does not imply specific recoverability or economic feasibility.

(ii) CDF and CCDF are complementary.

(iii) The complementarity lies in the relationship between the CDF and CCDF, not in the specific probabilities such as P90 and P10.

➡ In the context of the CDF, P90 indicates 10% probability of reserves are less than or equal to a given value, whereas in the context

of the CCDF, it indicates a 90% probability of reserves are greater than or equal to the same value.

➡ Similarly, P10 in the context of CDF represents 90% probability that the reserves are less than or equal to a value, whereas in the CCDF context, it represents 10% probability that reserves are greater than or equal to that value.

(iv) P50 in CDF and CCDF are equal.

➡ The P50 value (Median) represents 50% probability that the reserves are greater than or less than the given value.

(v) P90 and P10 reflect different confidence levels along the probability curves and are related to the shape of the distribution.

➡ They are not mathematically complementary.

(vi) P90, P50 and P10 do not correspond directly to P1, P2 and P3 (Proven, Probabable and Possible Reserves).

➡ P90, P50, and P10 are statistical measures used to describe the probability distribution of reserves, whereas P1, P2, and P3 refer to reserve classifications based on geological and economic certainty. However, there is a conceptual link.

(iv) These probabilistic terms reflect the range of uncertainty in resource estimates but are not directly linked to reserve classifications such as P1, P2, and P3.

2.5- P1, P2, P3 (Reserves):

These are categorical classifications used in petroleum resource estimation based on geological certainty and commercial viability:

➡ *P1 (Proved Reserves):* Reserves with high confidence of recovery. These are economically recoverable under current technology and market conuditions.

➡ *P2 (Probable Reserves):* Reserves with moderate confidence of recovery. There is a reasonable expectation that these reserves will be recoverable under favorable conditions.

➡ *P3 (Possible Reserves):* Reserves with low confidence of recovery. These depend on speculative factors such as future technology, market conditions, or additional discoveries.

2.6- 1P, 2P and 3P (Reserves):

1P, 2P, and 3P refer to different categories of reserves that reflect increasing levels of confidence in the geological and economic recoverability of the resource.

➡ ***1P Reserves*** **have the highest level of confidence in economically recoverablility. They are supported by strong geological, engineering, and economic evidence and are recoverable under current market conditions, technology, and regulations.**

➡ ***2P Reserves*** **are the sum of Proved Reserves (1P) and Probable Reserves P(2P). This category reflects a moderate level of confidence in recovery. Includes reserves that are less certain than 1P but still have a reasonable probability of being recovered.**

➡ ***3P Reserve*****s are the sum of Proved Reserves (1P), Probable reserves (P2) and Possible Reserves (P3). This category reflects a low level of confidence in recovery. It includes reserves that are more speculative and dependent on favourable future conditions (e.g. price increases, technological advances, regulatory changes).**

2.7- Statistical Descriptors

a) Minimum:

The lowest possible reserve estimate based on available data, typically reflecting pessimistic scenarios.

b) Maximum:

The highest possible reserve estimate based on available data, representing the upper bound of uncertainty.

c) Mean:

The average reserve estimate, calculated as a weighted average of all possible reserve outcomes. It provides a central tendency but is sensitive to extreme values. Represents the expected average reserve volume, taking into account all possible scenarios (including outliers). It is most applicable to symmetric distributions (e.g., normal), where the data is uniformly distributed around the central value

Mean = (sum of all values) / (umber of values)

Example:

- ***Reserve estimates (Mb)*****: 10, 20, 30, 40, 150**
- ***Mean*** **= (10 + 20 + 30 + 40 + 150) / 5 =**

Mean = 50 Mb

d) Median:

The middle value when the data set is ordered from smallest to largest. If the number of data points is odd, the median is the middle value. If the number of data points is even, the median is the average of the two Mean values.

e) Median (≈ P50):

Represents the "best estimate" where reserves are equally likely to be higher or lower, making it more robust to skewed distributions such as reserve data.

Example:

- **Reserve estimates (Mb): 10, 20, 30, 40, 150**
- **Ordered = 10, 20, 30, 40, 150**

Median = 30 MB

f) Mode:

The probable reserve estimate corresponding to the peak of the probability density function (PDF).

In probabilistic reserve evaluation, a lognormal distribution is often used when reserve data are positively skewed, reflecting the nature of subsurface uncertainties.

Feature	Mean	Median
Definition	Arithmetic average	Middle value in ordered data
Outliers Effect	Highly influenced by outliers	Not influenced by outliers
Best Use	Symmetrical (normal) distributions	Skewed distributions
Interpretation	Balance point of the data	Typical central value

The P90, P50, P10, Mean, Mode, Minimum, and Maximum values are calculated and interpreted for a lognormal distribution. For a lognormal distribution:

➡ **The Mean, Mode, and Median (P50) differ due to the skewness of the distribution.**

➡ **Percentiles (P90, P50, P10) are extracted from the Cumulative Distribution Function (CDF) of the lognormal curve.**

➡ **Main Differences between Mean and Median:**

3- Calculating Key Metrics for Reserves

Let's look at two examples of reserve key metrics calculations:

- **In example 1, we assume a triangular probability distribution for reserves.**

- **In example 2, we assume a lognormal probability distribution, which is often used when reserve estimates are positively skewed.**

3.1- Example 1: Triangular Distribution

Consider a triangular hypothetical probability distribution for reserves, characterized by the Probability Density Function (illustrated on the next page) and defined by the following parameters: (i) Minimum: 10 Mb, (ii) Mode: 25 Mb and (iii) Maximum: 50 Mb

3.1.1- Metric Calculations

Monte Carlo simulation or the Cumulative Distribution Function (CDF) allows us to easily perform the calculations:

a) ***P90:*** **Read at the 90th percentile of the cumulative probability curve (see figure below). In this example, it is:**

P90 = 15 Mb

b) ***P50:*** **Read at the 50th percentile (Median). In this example, is:**

P50 = 30 Mb

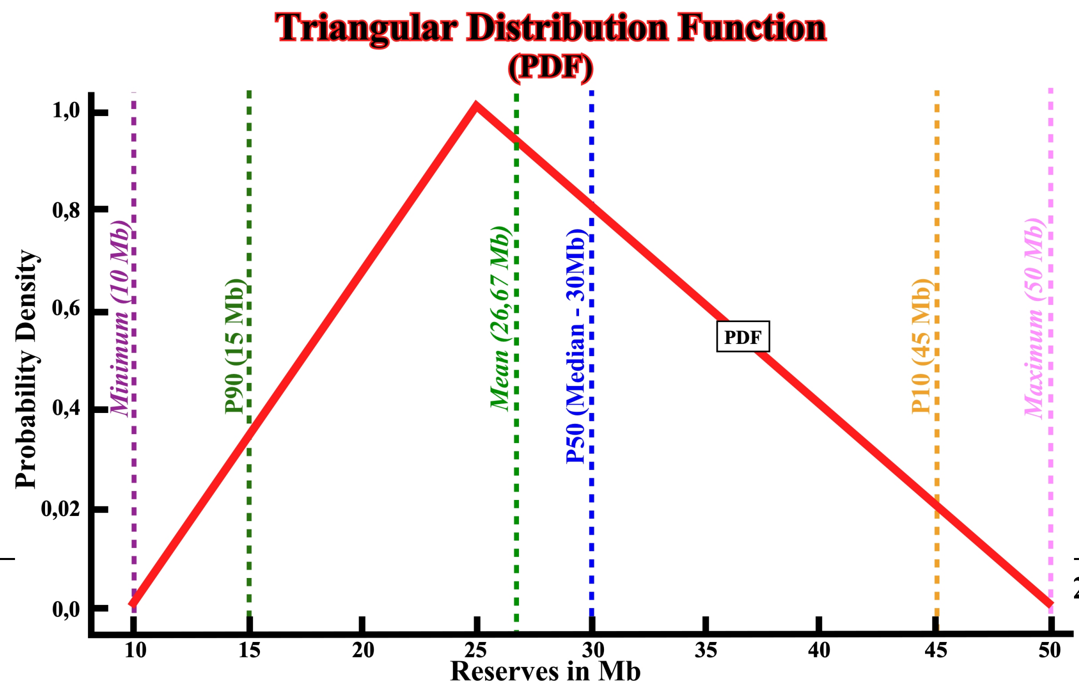
c) ***P10:*** **Read at the 10th percentile. This could be it:**

P10 = 45 Mb

d) ***Mean:*** **Calculated as [(Minimum + 4(*)) × Mode + Maximum], i.e. [(10+4) × 25 + 50] ÷ 6 , the average is :**

Mean= 26,67 Mb

(*) The 4 is a weight reflecting the shape of the triangular distribution, where the Mode is the most likely value and typically has the highest probability density. This weighting emphasises the importance of the Mode in the average, recognising that it is more likely than the Minimum and Maximum: (i) the Minimum and Maximum are each given a weight of 1, reflecting their contributions to the range of the distribution; (ii) the Mode is given a weight of 4, reflecting its centrality and higher probability within the triangular shape; (iii) the denominator 6 (1 + 4 + 1) normalises these weights to calculate the mean.



3.1.2- The triangular distribution of this example shows that:

- ☑ ***P90 (15 Mb)***

  is indicated by the green dashed line (conservative reserve estimate).

- ☑ ***P50 (Median - 30 Mb)***

  is indicated by the blue dashed line (the "best estimate" or Median.)

- ☑ ***P10 (45 Mb)***

  is indicated by the orange dashed line (optimistic reserve estimate=.

- ☑ ***Minimum (45 Mb)***

  is indicated by a purple dashed line.

- ☑ ***Mean (26,67 Mb)***

  is indicated by the light green dashed line.

- ☑ ***Maximum (50 Mb)***

  is indicated by the pink dashed line.

3.1.3- Key Insights on Reserve Metrics and Their Relationships

- ✳ **The percentiles (P90, P50, P10) indicate the probability of being less than or equal to a given reserve volume."**
- ✳ **The Mean indicates the expected value.**
- ✳ **The Mode emphasis the most likely outcome.**
- ✳ **The Minimum and Maximum reflect the range of possibilities.**

Metric	Value (Mb)	Description
Minimum	10	Lowest possible reserve estimate
Mode	25	Most likely reserve estimate
P90	15	Conservative estimate (90% chance)
P50	30	Best estimate (50% chance)
Mean	26,67	Average reserve estimate
P10	45	Optimistic estimate (10% chance)
Maximum	50	Highest possible reserve estimate

- **For skewed distributions, the Mean may differ significantly from the Median.**

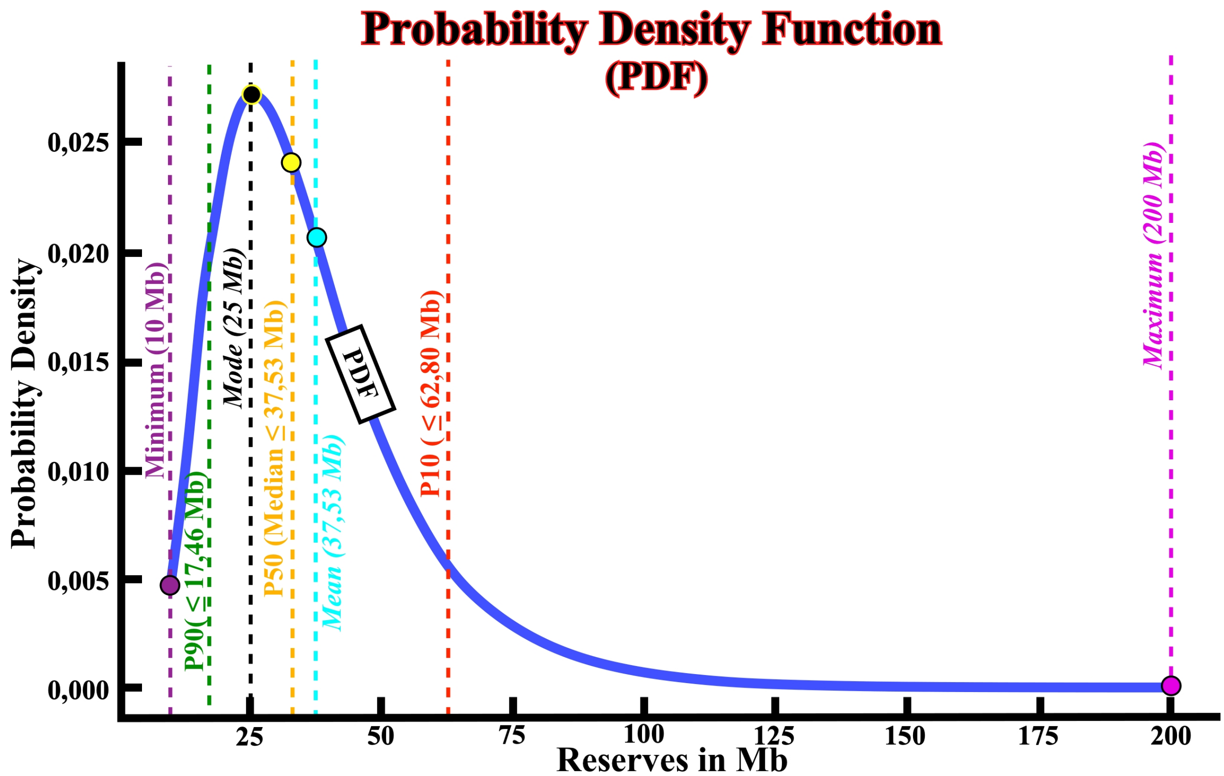
3.1.4- Summary Table of Hypothetical Results:

3.1.5- Recommendations

- **Use percentile metrics (P90, P50, P10) for decision making under uncertainty.**
- **Supplement statistical descriptors (Mean, Mode) with percentile estimates for full understanding.**
- **Ensure that input parameters and distributions reflect realistic uncertainties.**

3.2- Example 2: Lognormal Distribution

The graph below illustrates the Probability Density Function (PDF) used to estimate the reserves of a specific oil field.



Notice that:

(i) The peak represents the Mode, indicating the most likely reserve value

(ii) The vertical lines mark key reserve metrics:

➡ **P90 (green line):**

indicates 10% probability that reserves are

≤ 17,46 Mb.

➡ **P50 (Median, orange line):**

indicates 50% probability that reserves are

≤ or ≥ 33,53 Mb.

➡ **P10 (red line):**

indicates 90% probability that reserves are

≤ 62,80 Mb.

➡ **Maximum (pink line):**

reflects the upper limit ≈

≈ 200 Mb.

➡ **Minimum (purple line):**

reflects the assumed lower limit ≈

≈ 10 Mb.

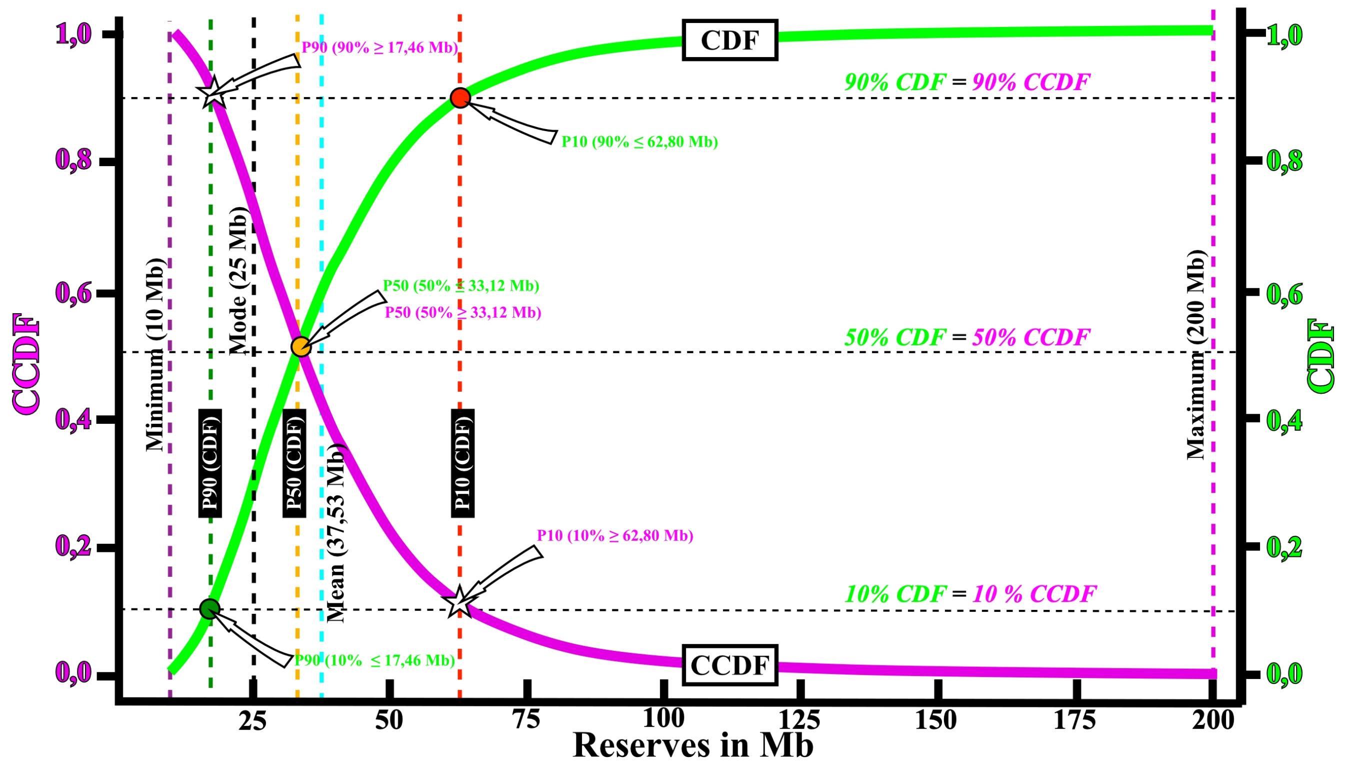
➡ **Mean (cyan line):**

calculated lognormally ≈

≈ 37,53 Mb.

(iii) In the graph below, the Cumulative Density Function (CDF) for the reserve evaluation is shownin green colour. It represents the probability that reserves are less than or equal to a given value.

Complementary Cumulative Distribution Function (CCDF)
Cumulative Distribution Function (CDF)



➡ **At P90, in CDF, there is 10% probability that reserves are less than or equal to to the associated reserve level.**

➡ **This leaves a 90% (1- 0,1 = 0,9) probability that reserves are greater than or equal to this reserve level, which is shown in the Complementary Cumulative Density Function (CCDF) shown in purple on the same graph.**

➡ **The CCDF, also known as Survival Function, indicates that at P90 there is a 90% Probability that the reserves are greater than or equal to such a reserve level.**

➡ **In the graph above, the horizontal lines represent key probabilities:**

☑ ***90% (or 0,9) probability in the CCDF:***

- At P90, there is a 10% probability that reserves (in the CDF) are less than or equal to 17,46 Mb and a 90% probability that reserves (in the CCDF) are greater than or equal to 17,46 Mb.

☑ ***50% (or 0,5) probability in the CCDF:***

- At P50, there is a 50% probability that reserves (in CCDF) are greater than or equal to 33,12 Mb and a 50% probability that reserves, in the CDF, are less than or equal to 33,12 Mb.

- P50 indicates a Median estimate, where reserves are equally likely to be greater than or less than 33,12 Mb.

☑ ***10% (or 0,1) probability in CCDF indicates that:***

- At P10, there is a 10% probability that the reserves will be greater than or equal to 62,80 Mb and 90% probability that reserves, in CDF are less than or equal to 62,80 Mb.

- P10 represents an optimistic scenario where there is a 10% probability that reserves will exceed 62,80 Mb.

Additional Notes:

CCDF vs. CDF Probabilities:

✳ **Probabilities for the CCDF are greater than or equal to, while probabilities on the CDF are less than or equal to.**

P90, P50 and P10 Scenarios:

✳ **P90 represents a conservative (low) reserve scenario, providing high confidence that the minimum reserve thresholds will be met.**

- ✳ **P50 is the "best estimate" or Median because, in a probabilistic distribution, the 50% line represents the value at which the distribution is evenly balanced.**

- ✳ **P10 represents an optimistic (high) reserve scenario, consistent with potential upside.**

➡ **In the graph above, the vertical lines represent reserve percentiles**

- ☑ **These lines correspond to the P90, P50 and P10 contingency values, indicating specific reserve amounts for the respective confidence levels.**

- ☑ **In addition, Minimum, Mean and Maximum are highlighted using a consistent colour code : Purple for the Minimum), Cyan for the Mean) and Pink for the Maximum).**

4- Summary of CDF Use and Benefits

In assessing reserve confidence levels and supporting risk-based decision making, the CDF graph is a valuable tool for decision makers:

Conservative Estimate (P90)

✤ **The P90 value represents a conservative estimate, indicating that there is a 90% probability that reserves will be less than or equal to a given value.**

✤ **It is typically used when a high degree of certainty is required, particularly in planning scenarios where minimising risk is a priority (e.g. securing financial commitments or planning for minimum reserves).**

Balanced Estimate (P50)

✤ **The P50 value corresponds to the median of the distribution, where there is an equal 50% probability that reserves will be either below or above the given value.**

✤ **This is the most likely or best estimate and is ideal for balanced decision making as it reflects the central tendency of the data without over-emphasising optimistic or conservative scenarios.**

Optimistic Estimate (P10)

✤ **The P10 value represents an optimistic scenario, indicating that there is only a 10% probability that reserves will be less than or equal to ta given value. This value is used to explore the potential upside in reserves, often to dis-cuss best case scenarios, future growth opportunities, or aggressive project strategies.**

Key benefits of using the CDF

✤ **The CDF provides clarity on risk and confidence levels, allowing decision makers to tailor their decisions to their tolerance for risk and uncertainty.**

- ✤ The ability to relate probabilities directly to reserve estimates allows quantitative comparisons between different projects or scenarios.
- ✤ CDFs can be combined with economic or operational models to refine resource development, investment and production decisions.

5- Understanding Log-Transformed Data and Calculating Parameters

5.a) Log-Transformed Data

- ➡ The log-transformed data refers to the values obtained by taking the natural logarithm (ln) of the original data.
- ➡ In the context of a log-normal distribution, the application this transformation tranformss the positively skewed log normal data into a symmetric normal distribution, there simplifying the analysis.

5.b) Types of Standard Deviation

When working with a log-normal distribution, two types of standard deviation need to be distinguished:

a) Standard Deviation of the Underlying Normal Distribution (σ):

- ➡ This is the standard deviation of the natural logarithm of the variable (ln(x)).

b) Standard Deviation of the Log-normal Distribution (SDx):

- ➡ This is the standard deviation of the original variable x, which follows the lognormal distribution.
- ➡ The Standard Deviation of the log-normal distribution (SDx) can be calculated using the relationship between the parameters of the log-transformed normal distribution (μ and σ) and the lognormal distribution:

$$SDx = M \times \sqrt{(e\sigma^2 - 1)}$$

which in our example will be SDx = 20,03 (see 5.g (vii)).

5.c) Log-Transformed Data Distribution:

- ➡ The natural logarithm of ln(x) follows a normal distribution:

$$\ln(x) \sim N(\mu,\sigma^2)$$

where:

μ = Mean of ln(x) (log-transformed data).
σ^2 = Variance of ln(x).
$\sigma = \sqrt{\sigma^2}$, the standard deviation.

5.d) Relationship Between the Mean (M) and the Parameters (μ and σ):

➡ **The mean of the lognormal distribution (M) is related to the log-transformed parameters by:**

$$M = e^{(\mu + \sigma 2/2)}$$

Take the natural logarithm:

$$\ln(M) = \mu + \sigma^2 / 2$$

thus:

$$\ln(M) = e^{(3,501 + 0,125)} =$$

$$= e^{3,626} =$$

$$M = 37,60 \text{ Mb}$$

Note: **ln(M)** ***refers to the natural logarithm of the mean of the lognormal distribution (M), derived from the relationship*** $M = e^{\mu + \sigma 2/2}$***. This is different from*** **μ*****, which is the mean of the log-transformed data.***

Rearrange to find σ^2 **:**

$$\sigma^2 = 2 \times \ln(M) - \mu$$

5.e) Calculate μ and σ:

➡ **To calculate both μ and σ, a second independent piece of information is required, such as:**

- ✳ **The standard deviation of the lognormal distribution (SDx).**
- ✳ **Percentile values (e.g., P90 and P10).**

5.f) Using Percentiles P90 (17,46 Mb) and P10 (62,80 Mb)

This section explains how to calculate the mean (μ) and standard deviation (σ) of the log-transformed data (ln(x)) from the P90 and P10 percentiles of a lognormal distribution.

a) Relation of Percentiles to μ and σ :

The lognormal distribution has the property that its percentiles can be expressed as:

$$P90 = e^{\mu - 1,2816 \times \sigma} \quad \text{and} \quad P10 = e^{\mu + 1,2816 \times \sigma}$$

- **P90: This is the value below which 90% of the data lies. The factor 1,2816 is the Z-score for the 90th percentile in the standard normal distribution.**
- **P10: This is the value below which 10% of the data lies. Similarly, the factor 12816 is the Z-score for the 10th percentile.**

b) Take the Natural Logarithm:

To linearise the exponential expressions, we take the natural logarithm of both equations:

(i) For P90:

$$\ln(P90) = \mu - 1{,}2816 \times \sigma$$

(ii) For P10:

$$\ln(P10) = \mu + 1{,}2816 \times \sigma$$

These equations relate the log-transformed percentiles ln(P90) and ln(P10)) μ and σ.

c) Solve for μ and σ:

To find μ and σ, we solve these two equations:

➡ **Add to find μ:**

Adding the two equations calculates the σ term

$$\ln(P90) + \ln(P10) =$$
$$= (\mu - 1{,}2816 \times \sigma) + (\mu + 1{,}2816 \times \sigma) =$$
$$\{\ln(P90) + \ln(P10)\} = 2 \times \mu$$

Rearrangefor μ:

$$\mu = [\ln(90) + \ln(p10)] \div 2$$

➡ **Subtract the equations to find σ:**

Subtracting the first equation from the second isolates σ:

$$\{\ln(P10) - \ln(P90)\} =$$

$$(\mu + 1{,}2816 \times \sigma) - (\mu - 1{,}2816 \times \sigma)$$

Simplify:

$$\ln(P10) - \ln(P90) = 2 \times 1{,}2816 \times \sigma$$

Rearrange for σ:

$$\sigma = [\ln(P10) - \ln(P90)] \div (2 \times 1{,}2816)$$

➡ **Calculate the Variance (σ²):**

Once σ is known, the variance is simply the square of the standard deviation:

$$\sigma^2 = [\{\ln(P10) - \ln(P90) \div (2 \times 1{,}2816)]^2$$

➡ **Examples with given alues:**

✳Inputs:

P90 = 17,46 Mb and P10 = 62,80 Mb

✳Step 1: Compute ln(P90) and ln(P10):

The values of P90 (17,46Mb) and P10 (62,80Mb) were obtained from the Complementary Cumulative Distribution Function (CCDF). The natural logarithm of these values was taken:

$$\ln(P90) = \ln(17{,}46) \approx 2{,}859$$
$$\ln(P10) = \ln(62{,}80) \approx 4{,}142$$

These logarithmic values are used as inputs for further calculations of μ and σ.

✳Step 2: Calculate μ:

$$\mu = [\ln(P90) + \ln(P10)] \div 2 =$$
$$= (2{,}859 + 4{,}142) \div 2 \approx$$
$$\mu \approx 3{,}501$$

✳Step 3: Calculate σ:

$$\sigma = [\ln(P10) - \ln(P90)] \div (2 \times 1{,}2816) =$$
$$= (4{,}142 - 2{,}859) \div (2 \times 1{,}2816) =$$
$$= 1{,}283 \div 2{,}5632 = 0{,}500 =$$
$$\sigma = 0{,}500$$

✳Step 4: Calculate σ²:

$$\sigma^2 = (0{,}500)^2 = 0{,}250$$

5.g) The results of other calculations are:

(i) T*he Mean (M) of the lognormal distribution is*

$$M = 37{,}60 \text{ Mb}$$

(ii) *The Median (P50) is calculated as:*

$$\text{Median} = e^{\mu} =$$
$$= e^{3{,}501} \approx 33{,}12$$
$$\text{Median} \approx 33{,}12 \text{ Mb}$$

(iii) *The Mean (μ) of log-transformed data*

The relationship between the mean of the lognormal distribution (M) and the mean of the log-transformed data (μ) is given by:

$$\ln(M) = \mu + (\sigma^2 \div 2)$$

From early calculations:

$$\ln(M) = 3{,}626, \text{ and } (\sigma^2 \div 2) = 0{,}125$$

Rearrange to solve for μ:

$$\mu = \ln(M) - (\sigma^2 \div 2)$$

$$\mu = 3{,}626 - 0{,}125 = 3{,}501$$

$$\mu \approx 3{,}501$$

To check the calculation, using $\sigma^2 = (0{,}500)^2 = 0{,}25$, we calculate

$$\mu = \ln(M) - (\sigma^2 \div 2) =$$

$$= 3{,}626 - 0{,}125 \approx 3{,}501$$

$$\mu \approx 3{,}501$$

(iv) Standard Deviation (σ) of log-transformed data) =

$$\sigma = 0{,}5 \text{ Mb}$$

(v) Standard Deviation of the Lognormal Distribution (SDx):

Use the formula:

$$SDx = M \times \sqrt{(e^{\sigma^2} - 1)}$$

Replace the values:

$M = 37{,}60$ Mb,

$\sigma^2 = 0{,}25$

Calculate:

1) **Compute e^{σ^2}:**

$$e^{0{,}25} \approx 1{,}284$$

2) **Subtract 1:**

$$e^{\sigma^2} - 1 = 1{,}284 - 1 = 0{,}284$$

3) **Take the square root:**

$$\sqrt{0{,}284} \approx 0{,}533$$

4) **TMultiply by M:**

$$SDx = 37{,}60 \times 0{,}533 \approx 20{,}03 \text{ Mb}$$

$$SDx = 20{,}03 \text{ Mb}$$

(vi) Back-transformed Minimum and Maximum

Minimum = 10 Mb.   Maxi = 200 Mb

(vii) Other Parameters

➡ **(v) Mode:**

$$\text{Mode} = e^{(\mu - \sigma 2)} =$$

$$= e^{(3{,}501 - 0{,}025)} = e^{3{,}3376} =$$

Mode ≈ 29,12Mb

➡ **Mean:**

$$\text{Mean} = e^{[\mu+(\sigma 2 \div 2)]}$$

$$= e^{3,501+(0,125]} = e^{3,62} =$$

Mean ≈ 37,60 Mb

➡ **Median:**

$$\text{Median} = e^{\mu} =$$

$$= e^{3,501} =$$

Median ≈ 33,12Mb

Using the Using Cumulative Distribution Function of the lognormal P90 and P10 percentiles can be calculated numerically, knowing that

a)The CDF of the lognormal distribution is:

$$F(x)=\Phi\ [(\ln(x) - \mu) \div \sigma]$$

where

Φ(z) is the CDF of the standard normal distribution.

b) P90 is the value of x such that F(x) = 0,9.

c) P10 is the value of x such as F(x) = 0,10

Metric	Value (Mb)	Description
Minimum	10	Assumed lowest reserve estimate
Mode	29,12	Most likely reserve estimate
P90	17,46	
P50 (Median)	33,12	Conservative estimate (90% chance)
Mean (M)	37,60	Best estimate (50% chance)
P10	62,80	Optimistic estimate (10% chance)
Maximum	200	Assumed highest reserve estimate
μ	3,50	Mean of the log-transformed data (ln(x))
σ	0,5	Standard Deviation of Log-transformed
σ^2	0,25	Variance of the log-transformed data
SDx	20,33	Standard Deviation of the original lognormal data

Using a numerical root-finding methods (e.g. Newton-Raphson, bisection, or secant methods) we can easily find it:

P10: 17,46 Mb (as expected)

P90: 62,89 Mb (as expected)

In this chapter, we have demonstrate the use of lognormal distribution analysis to calculate key metrics, including the Mean, Median, Mode, and Standard Deviations, and to validate validate percentiles. These calculations provide key insights for reserve estimation in petroleum exploration and can be applied to other data sets that exhibit lognormal behaviour.

6- Conclusions

6.1- Understanding Uncertainty:

- ★ **P90 provides high confidence for conservative planning, provinding a secure baseline.**
- ★ **P50 provides a balanced median estimate, ideal for moderate risk decisions.**
- ★ **P10 highlights potential upside, helping to explore growth opportunities.**

6.2- Understand Reserves:

- ★ **The Mean reflects the expected value but is skewed higher in positively skewed distributions.**
- ★ **The Median (P50) provides a robust central estimate unaffected by extreme values.**
- ★ **The Mode indicates the most likely reserve value, emphasising realistic outcomes.**
- ★ **Minimum and Maximum values define the boundaries of reserve estimation, shaping the range of possibilities.**

6.3- Practical Recommendations:

- ★ **Use probabilistic metrics such as P90, P50, and P10 for tailored decision-making.**
- ★ **Validate input assumptions to ensure realistic reserve estimates and effective risk management.**

By integrating probabilistic insights with visual tools (e.g., PDFs and CDFs), decision makers can achieve a comprehensive understanding of reserves and associated risks. This increases confidence in resource planning and investment strategies.

Appendix

(Major Takeaways)

1. Definition of Percentiles:

- **A percentile is a statistical measure that indicates the relative position of a value within a data set or distribution. The n^{th} percentile (Pn) of a data set is the value below which n% of the data lies. For example, the 90^{th} percentile (P90) means that 90% of the values in the data set are less than or equal to the associated reserve level.**
- **However, in the oil and gas industry, particularly in reserve evaluation, decision makers often interpret P90 in terms of the Complementary Cumulative Distribution Function (CCDF) rather than the standard Cumulative Distribution Function (CDF). This approach emphasises the probability of exceeding the reserve level.**
- **For example, if P90 = 17.48 Mb, the CDF probability indicates that 90% of reserves are less than or equal to 17.48 Mb. However, decision makers focus on its complement, the 90% probability that reserves exceed this value.**
- **This interpretation aligns P90 with conservative planning, providing high confidence that reserves will exceed a minimum threshold. As a result, P90 is widely used to backstop financial commitments or establish baseline reserves.**

2. Percentile in Reserve Evaluation

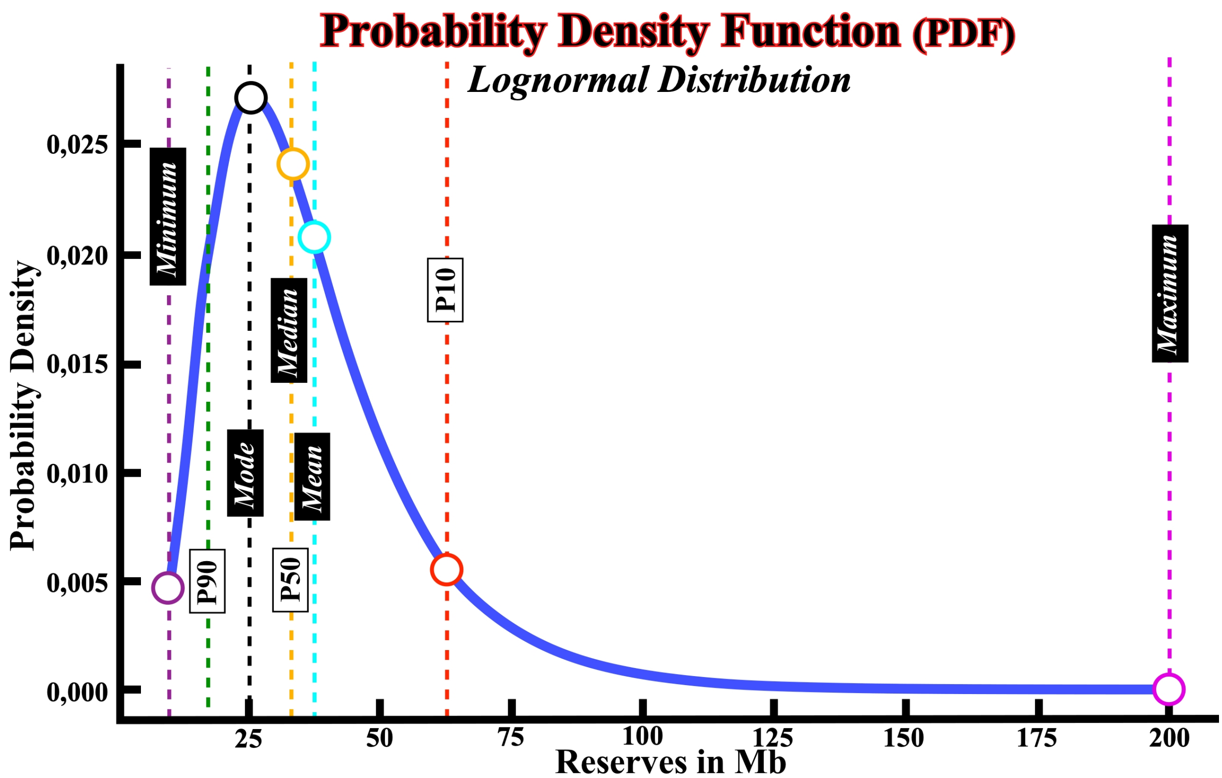
- **P90 (90th Percentile):**

  ➡ **90% of the values are below or equal, while 10% are above.**
- **P50 (50th Percentile or Median):**

  ➡ **50% of the values are below or equal, and 50% are above.**
- **P10 (10th Percentile):**
  - **10% of the values are below or equal, and 90% are above**

3. Percentiles (P90, P50, P10):

- **P90 (Proved Reserves): Represents a conservative estimate that provides 90% confidence that reserves will exceed the associated reserve level. It is generally closer to the Minimum in skewed distributions.**
- **P50 (Median, Probable Reserves):**
  Represents the best estimate or central value. It lies between P90 and P10 and is often close to the Mode in symmetrical distributions. In skewed distributions such as log-normal, P50 is higher than the Mode.
- **P10 (Possible Reserves):**
  An optimistic estimate, with only a 10% probability of reserves exceeding he associated reserve level. This value is closer to the Maximum in positively skewed distributions.



4. Mean vs. Median:

- **Mean (Expected Value):**
  The Mean provides the arithmetic average of the reserve estimates and takes into account the entire distribution. In positively skewed distributions (e.g., lognormal), the Mean is always higher than the Median due to the influence of the long tail.

- **Median (P50):**
  Being the middle value, the Median is not influenced by the skewness of the distribution and offers a robust central estimate.

5. Mode vs Mean:

- **Mode:**
  Represents the most likely value of the reserves. For positively skewed distributions, the Mode is the lowest of the three central tendencies (Mean, Median, Mode). In symmetric or near-symmetric distributions (e.g., triangular), It is close to the P50.

- **Mean vs. Mode:**
  The difference between the Mean and Mode increases with the degree of skewness of the distribution. In highly skewed distributions, the mean moves further towards the maximum.

- **Minimum:**
  Reflects the lowest plausible reserve estimate, defining the lower limit of the distribution. In practice, it often correlates closely with P90 in narrow distributions..

- **Maximum:**
  Reflects the highest plausible reserve estimate, defining the upper limit of the distribution. In practice it often correlates with P10.

6. General Relationships:

- **P90, P50, and P10 provide a probabilistic framework for decision making, with P90 being the most conservative and P10 being the most optimistic.**
- **The Mean often provides a higher estimate than the median in skewed distributions, while the mode remains the lowest among the central values.**
- **Minimum and Maximum values are conceptual boundaries, and are less likely to correspond directly to a probabilistic percentile unless the range of uncertainty is very narrow.**

7. Visual Insights:

- **The PDF and CDF plots help to illustrate these relationships:**
  - **The PDF shows the location of the mode, mean, and maximum probability density.**
  - **The CCDF highlights the probability thresholds for P90, P50, and P10 and their relative positions within the distribution.**

8. Practical Recommendations:

1. **Use P90, P50, and P10 for communication and planning under uncertainty.**
2. **Use the mean for decision making where risk tolerance is moderate.**
3. **Be aware of the differences between Mean, Median, and Mode, especially in skewed distributions, to avoid overestimation or underestimation.**
4. **Validate assumptions for minimum and maximum as they strongly influence the shape and spread of the probabilisty distribution.**