Uncertainty Distributions

In practice

? Which – How – What ?

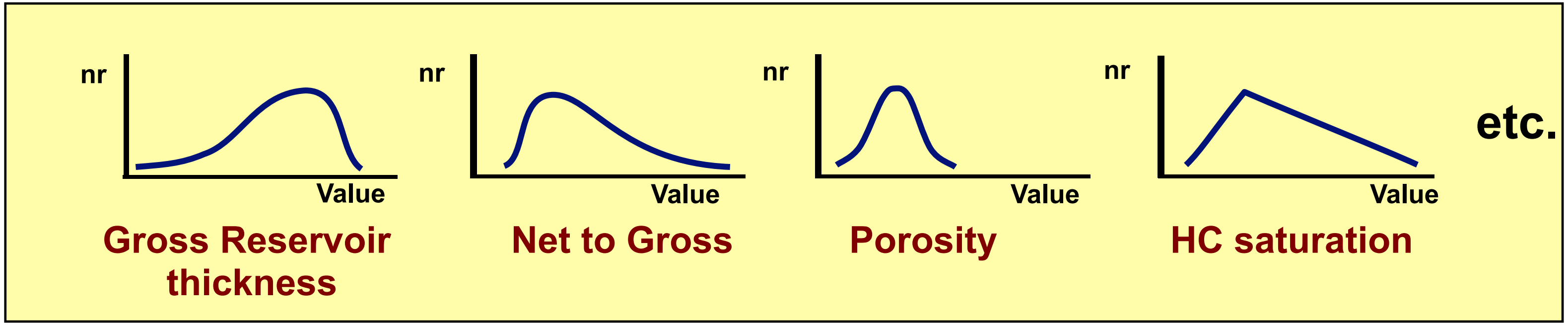
Jan de Jager

Exploration Director

petroEDGE

Probabilistic tools require uncertainty distributions

$$\text{Area} \times \text{Thickness} \times \text{N/G} \times \text{Ø} \times \text{Sat}_{hc} \times \text{FvF} \times \text{RF}$$

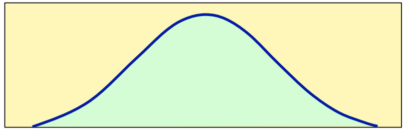


- Probabilistic tools have several distribution types to choose from
  - E.g. GeoX, REP, Prospect Risk Analysis Suite, Crystal Ball, …
- Are there preferred distribution types?
- Is this about statistics?

petroEDGE

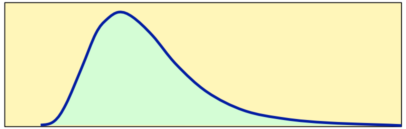
The most common Distribution Types

Normal



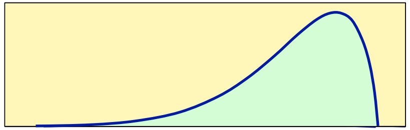
Symmetrical distribution

Lognormal



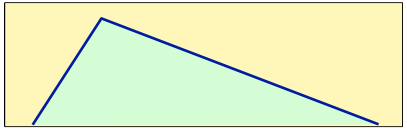
Asymmetrical distribution;
Positive skew

Beta



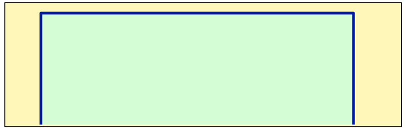
Distribution symmetrical or asymmetrical
Skew positive or negative

Triangular



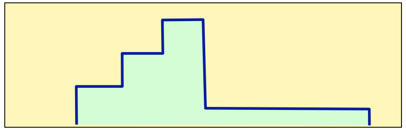
Distribution symmetrical or asymmetrical
Skew positive or negative

Uniform



Same probability for all values in range

Histogram



Custom ‘hand-made’ distribution

Lognormal Uncertainty Distributions

- There is a widespread *(but unjustified)* believe that for most input parameters a lognormal distribution type is preferred.
- The reasons normally come from:
  - *Megill, 1984; Capen 1984 & 1992:*
    - Most geotechnical parameters involved with oil and gas occurrences are lognormal
  - *Rose, 1996 & 2001:*
    - "Geoscientists who are aware of the prevalence of lognormality (and who constrain their estimates in the expectation of lognormality) will tend to make better predictions of most parameters having to do with oil and gas reserves"
      - "A few petroleum parameters are exponential; fewer still are normal"
    - "Predictions of all such parameters should be constrained by the expected form of the distribution"

P.R. Rose (2001): Risk Analysis and management of Petroleum Exploration Ventures. AAPG Methods in Exploration Series, No 12.

petroEDGE

Why are Distributions Confusing?

- **It seems that selection of distribution types is about statistics**

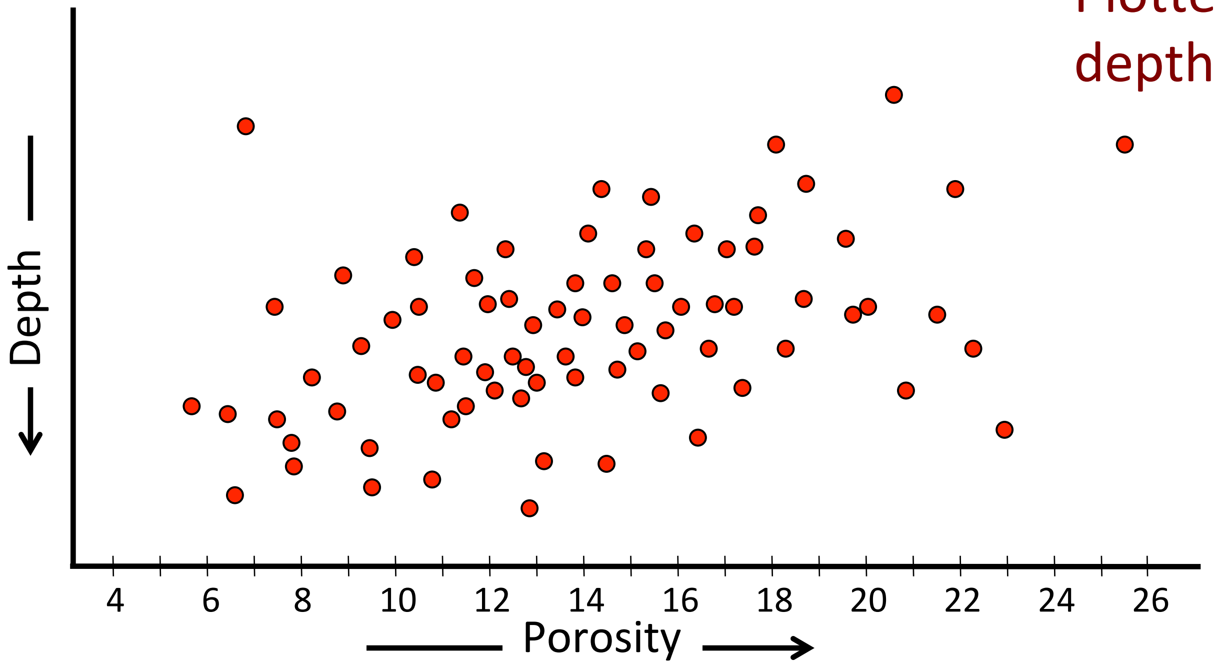
- **Distributions are used for:**
  - **Variability**
    - Mathematical description of the spread of data points
      - E.g. All porosities measured in a well or in a reservoir interval
  - **Uncertainty**
    - Indicating our uncertainty of the precise value of a parameter, such as e.g. the average porosity of a reservoir interval
      - Which depends on data, models and understanding

Based on: S. H Begg, R.B. Bratvold & M.B. Welsh (2014):
Uncertainty vs Variability: What's the difference and why is it important. SPE 169850.

Variability

- **Porosities measured in a reservoir interval in a well**

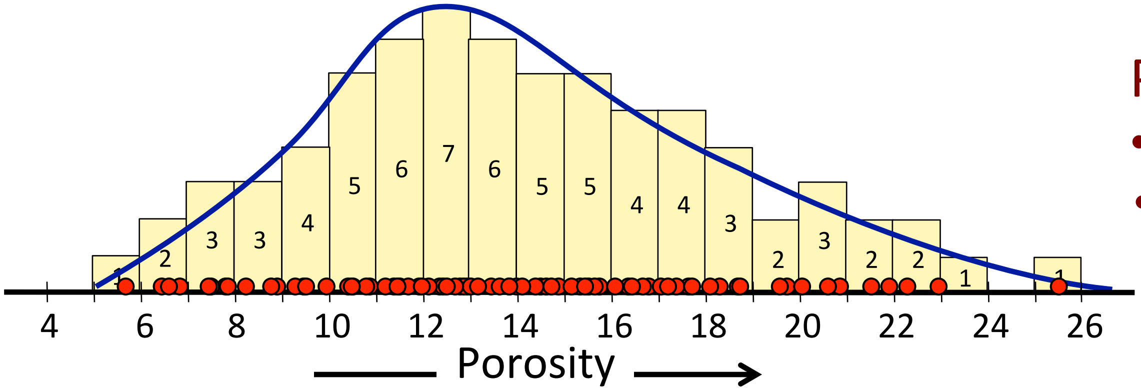
Plotted versus depth



petroEDGE

Variability Example

- **All porosities measured in a reservoir interval in a well**



Plotted as:

- Histogramme
- Continuous ditribution

- The variability of many subsurface parameters may (seem) Lognormal
- That's why many explorers think that input for probabilistic volume calculations also needs to be Lognormal
- **Is that correct?**

petroEDGE

Uncertainty Distributions

- Uncertainty can be described with a distribution type
  - Just like variability…
- But, with uncertainty distributions we are not describing real (existing) distributions
- The actual average values for Porosity, Net/Gross, Sat_{HC}, Column length, etc, are **precise** values
  - Spiked values
- We are just indicating our uncertainty of the actual value
  - Such as (for example) the average porosity of a reservoir interval
- **Variability ≠ Uncertainty**

Uncertainty

- **The uncertainty is in our head**
  - We don't know what the precise values are
  - The **mode** is our best guess
    - Most likely value
  - The **range** is defined by:
    - Minimum and Maximum value of the parameters: P99 - P1
    - Or 'reasonable' low and high values: P90 - P10
- **The uncertainty depends on the data and understanding**
  - If these increase, the uncertainty become less, and the distribution changes
- **Therefore:** **- There is no 'correct' or 'wrong' distribution**
  - There is no prescribed or required distribution type
- **We just need to make sure the distribution is geologically reasonable**
  - **Based on data <u>and</u> geological understanding**

petroEDGE

What if we use Lognormal ?

- Let's say, for reservoir thickness input we have selected a lognormal distribution
- Required inputs:

	Minimum	–	**Mean**	–	**Maximum**
Or,	**P90**	–	**Mean**	–	**P10**

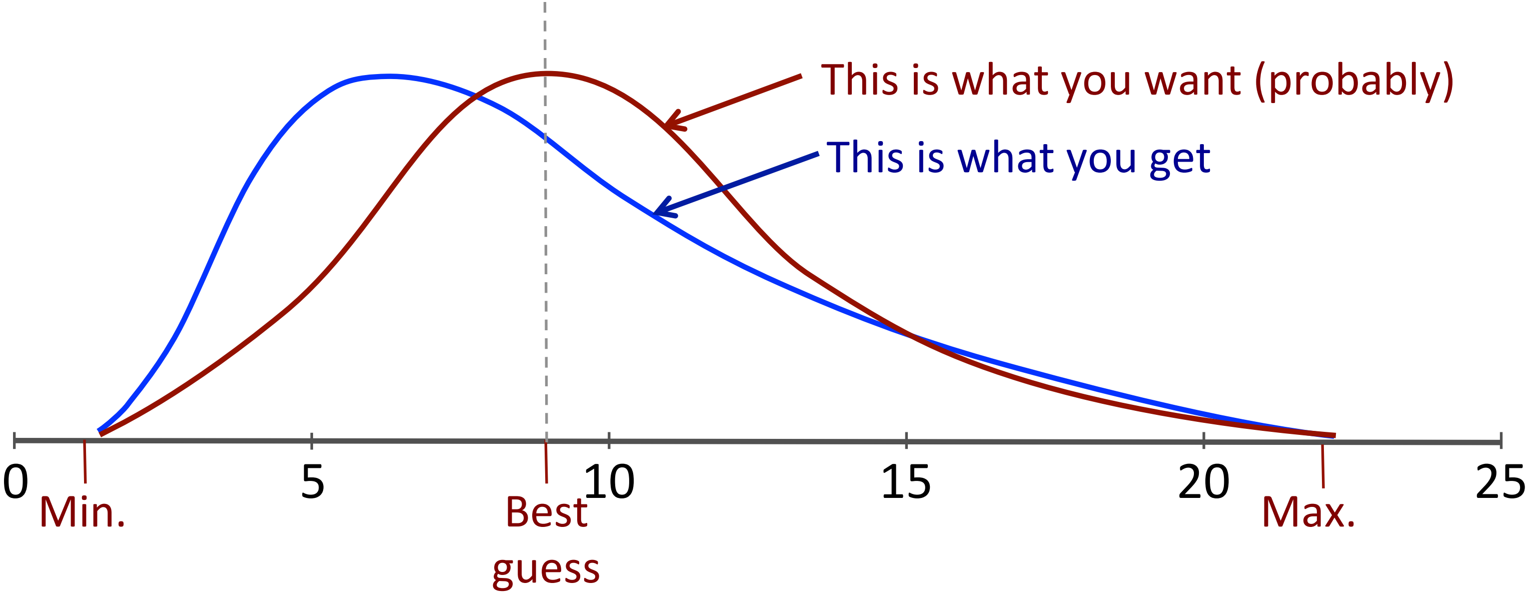
Which value would you choose for the Mean?

- ➢ **Your best guess for reservoir thickness?**
- ➢ **A higher value?**
- ➢ **A lower value?**
- ➢ **Something else?**

Input for reservoir thickness

For example: lognormal distribution

Minimum:	1 m
Best guess:	9 m
Maximum:	22 m

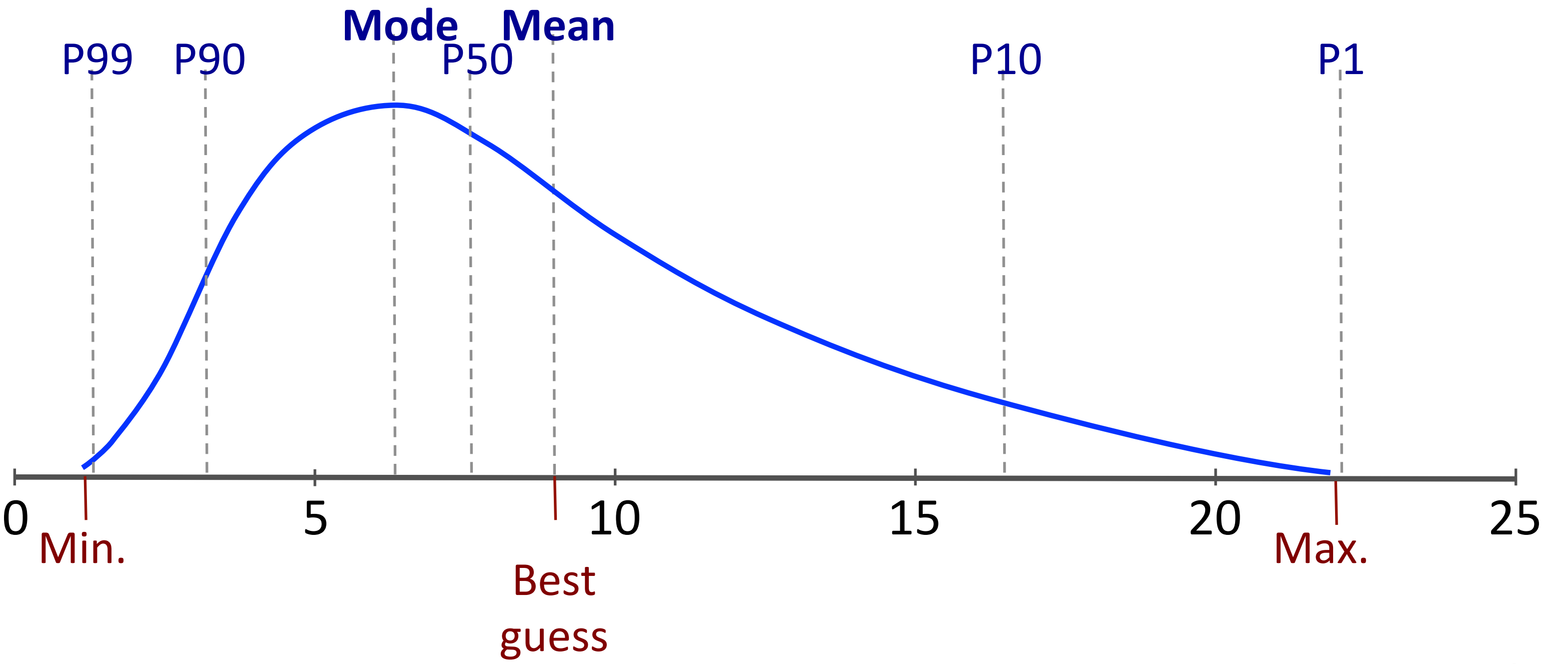


Input for reservoir thickness

For example:

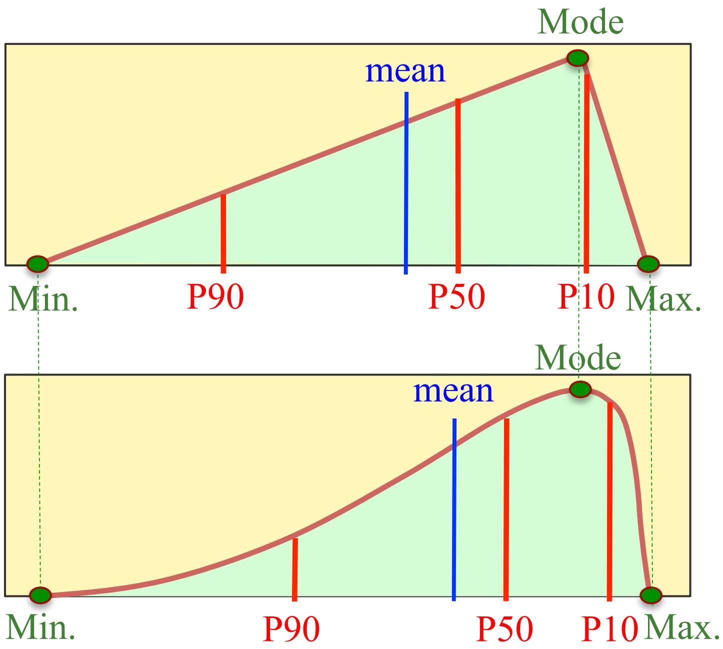
Minimum:	1 m
Best guess:	9 m
Maximum:	22 m

Mode = Most likely *by definition*



petroEDGE

Distribution Basics (1)

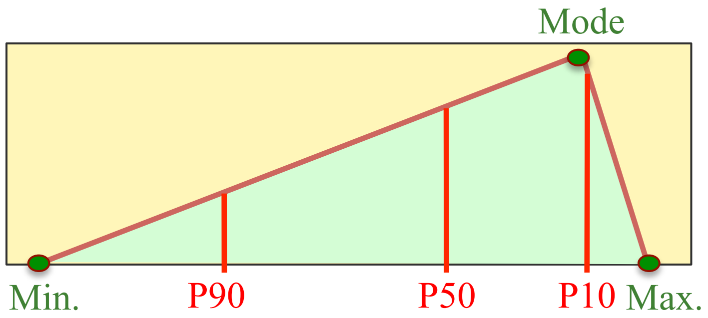


- In different distribution types, using the same input for Mode, Minimum and Maximum values, the P90, P50 and P10 values will be different

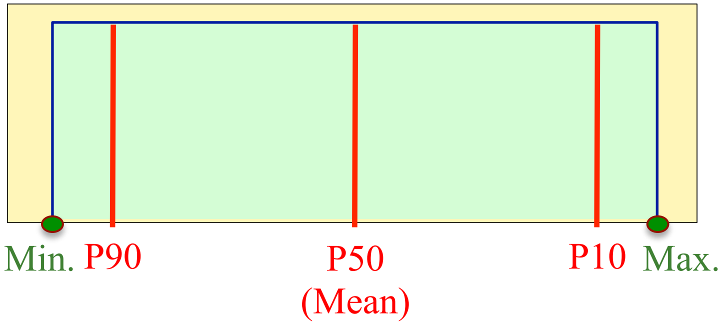
- In general, the P10/P90 ratio is maximised in Triangular distributions
  - Implying more uncertainty
  - Greater P10/P90

petroEDGE

Distribution Basics (2)



Uniform distrubution

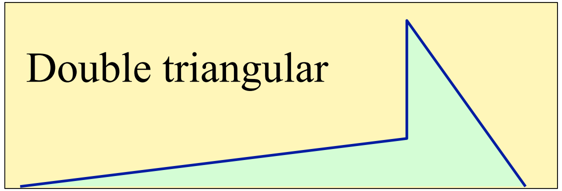


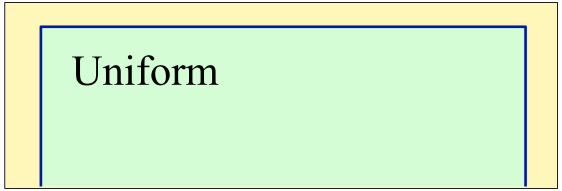
- **In a uniform distribution the P10/P90 ratio is even higher**
- **There is no Mode (most likely)**
  - The modal range is from minimum to maximum
- **There are abrupt differences in probability at the extremes**

petroEDGE

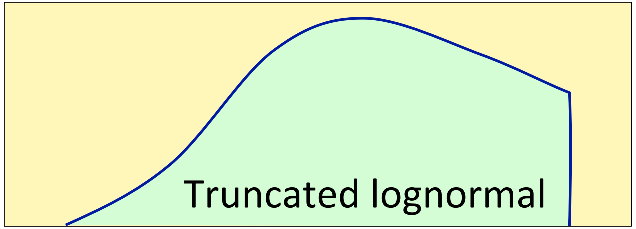
Pitfalls and Warnings

- Uniform or Double Triangular distributions are normally best avoided
  - These imply unlikely abrupt changes in probabilities





- **Avoid abrupt changes in probability**
  - These also occur with <u>Truncated</u> Normal and LogNormal distributions



Truncations should only be used at natural extremes, such as:

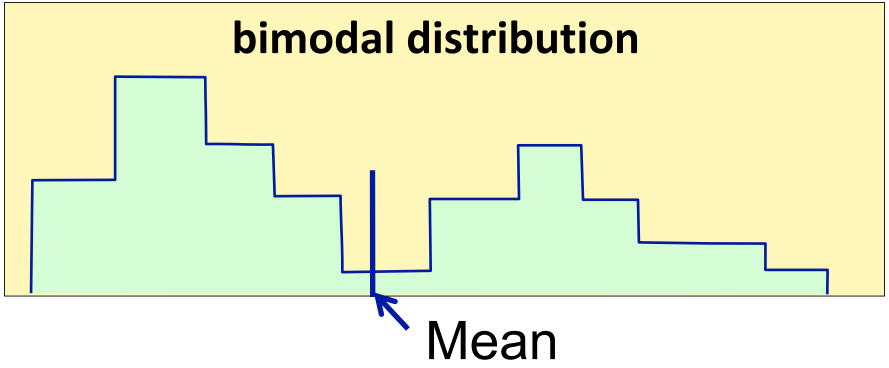
- 100% Net/gross
- 100% HC Saturation

petroEDGE

Pitfalls and Warnings (2)

Avoid bimodal distributions

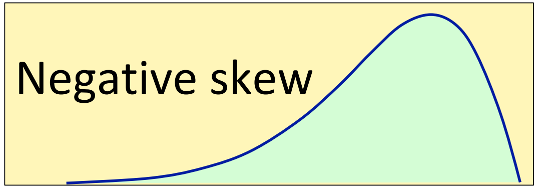
- The Mean may be a very unlikely value

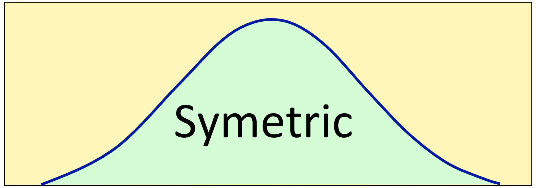


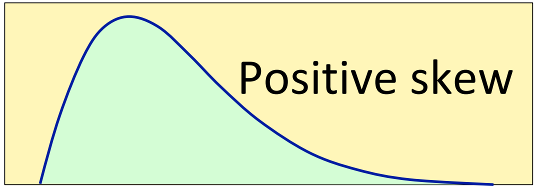
In case of bimodality it is better to use a scenario approach

Recommendations (2)

- Because of its versatility, the beta distribution can in most cases be used

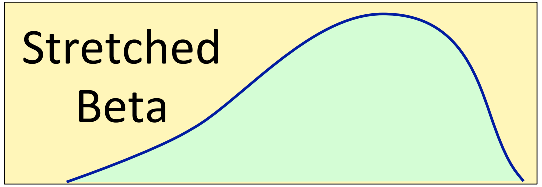






- Beta distribution is defined by:
  - Mode (= most likely)
  - Minimum and Maximum
  - Or P90 and P10

We should always be able to choose most likely (mode) and extreme values

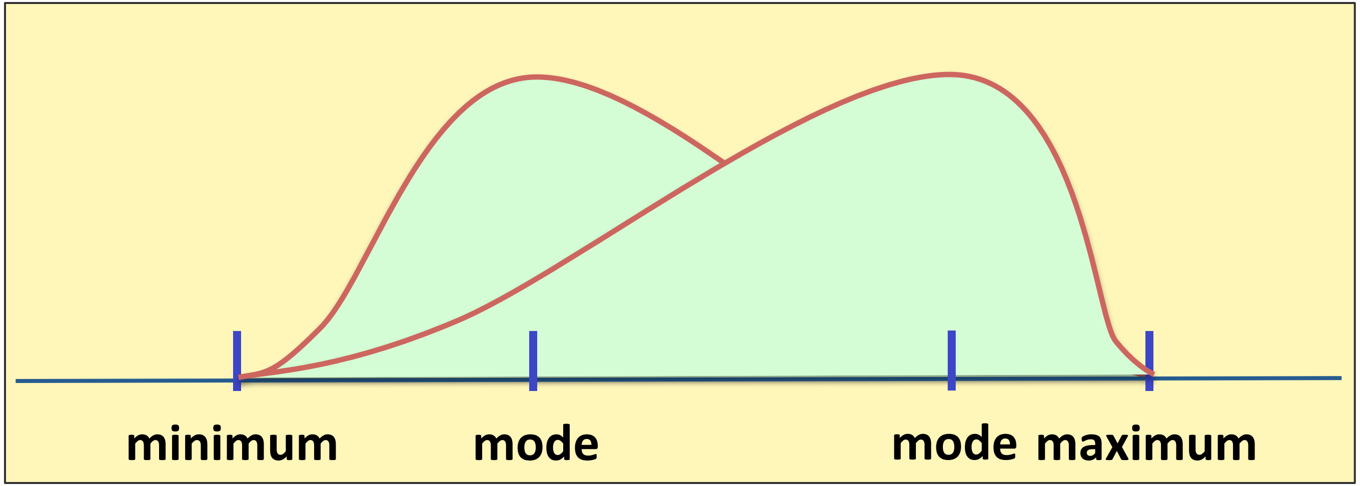


A stretched Beta distribution has a broader modal range, implying greater uncertainty

The Beta distribution as default allows us to focus on the geological uncertainty

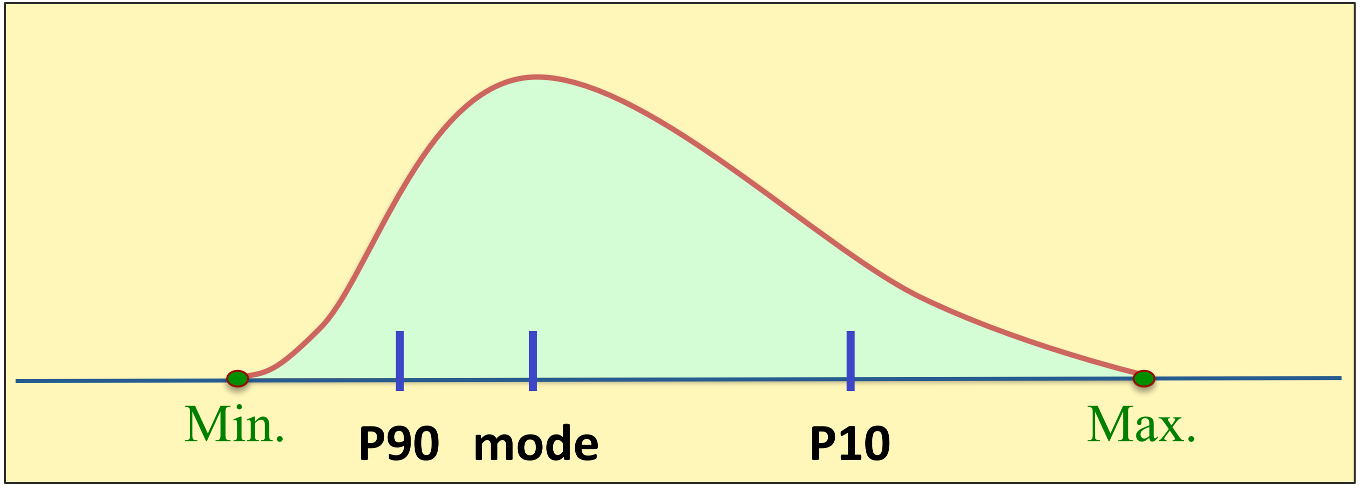
Drawing a Distribution

- Using the versatile Beta Distribution as a default allows focussing on the geological uncertainty



Drawing a Distribution

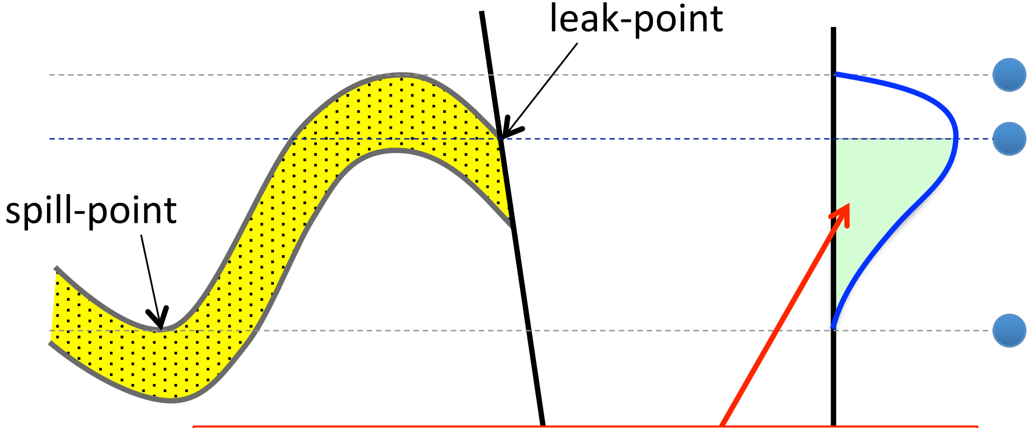
- The range can be indicated with Minimum and Maximum, or with P90 and P10



When using P90 and P10,
check that the extremes are geologically realistic

Beta is not always suitable

- A prospect with an independent dip-closure above a potential leak-point at a fault
- It is most likely that the trap will leak at the leak-point
- There is a small chance (*but unlikely*) that the fault may seal (partly)
  - giving a longer column, possibly to the deepest spill-point



Column length

Minimum: just below the crest

Mode: at the leak-point

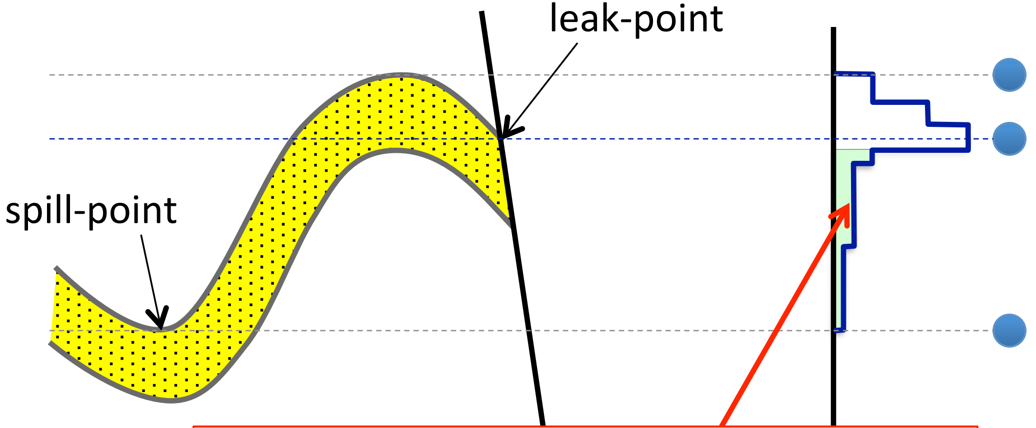
Maximum: at the spill-point

A Beta distribution can give a too high probability for a column longer than to the mode; in this case > 50%

Inconsistent with "unlikely"

Histogram may work better

- A prospect with an independent dip-closure above a potential leak-point at a fault
- It is most likely that the trap will leak at the leak-point
- There is a small chance (*but unlikely*) that the fault seals (partly)
  - giving a longer column, possibly to the deepest spill-point



Column length

Minimum: just below the crest

Mode: at the leak-point

Maximum: at the spill-point

With a histogram this can be adjusted to just a lower probability for a longer column

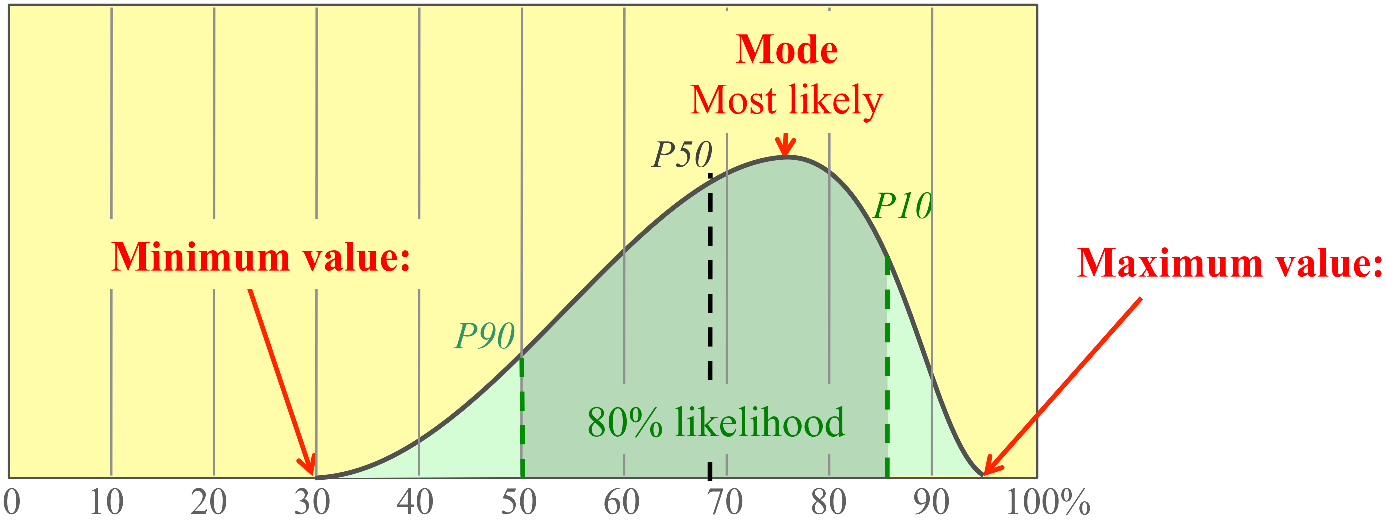
Consistent with "unlikely"

The Range

Area × Thickness × N/G × Ø × Sat_{hc} × FvF × RF

Uncertainty ranges must include

1. all outcomes that are geologically possible, and
2. that would be considered a technical success



petroEDGE

Conclusions

- **Uncertainty Distributions – *It's only geology !***
- **Don't be confused by statistics or probabilistic tools**
- **The Mode is your best guess (= most likely)**
- **Be realistic about the range**
  - Don't underestimate the uncertainties
  - Could it end up being a small sub-economic discovery?
  - Ask yourself how surprised you would be if the actual value falls outside the specified uncertainty range
  - If you would not be very surprised → increase the uncertainty range
- **The Beta distribution works well in most cases**
  - Sometimes a histogram is better