# Uncertainty Quantification

and Reliability Analysis

Henrique Kroetz

January 6, 2026

# The Inverse Problem in UQ

We wish to propagate uncertainty through a model, but first, we must characterize the uncertainty of the input parameters.

- We rarely know the "true" probability density function (PDF), $f_X(x)$.
- We must construct a PDF based on available information.

**Two Main Scenarios:**

1. **Sparse Data:** We rely on *Expert Information* and physical constraints.
2. **Rich Data:** We rely on statistical inference.

# Incorporating Expert Information

What if we have no data points, or very few? Standard statistical fitting is unreliable.

We must rely on **Expert Judgment** or **Physical Constraints**.

- **Bounds:** Is the parameter strictly positive? (e.g., stiffness, density). Is it bounded in $[0, 1]$? (e.g., porosity).
- **Moments:** Does the expert know the mean or variance?
- **Modality:** Is the value expected to cluster around a central point?

# The Principle of Maximum Entropy (MaxEnt)

We often have limited information (e.g., bounds, a mean value) but need to specify a full PDF $f(x)$.

- Choosing any distribution other than the one with maximum entropy amounts to assuming information we do not possess.
- MaxEnt provides the least biased estimate possible given the data.

## Shannon Entropy (Uncertainty Measure)

- Quantifies the average uncertainty or "surprise" in a random variable's outcomes.

$$H(f) = -\int_{\mathcal{D}} f(x) \ln(f(x))\, dx$$

We seek $f(x)$ that maximizes $H(f)$ subject to our constraints.

# Constructing the MaxEnt Distribution

We formulate this as a constrained optimization problem.

Maximize $H(f) = -\int f(x) \ln f(x)\, dx$.

Constraints:

- **Normalization:** Total probability must be 1.

$$\int f(x)\, dx = 1$$

- **Moment Constraints (Data):** The model averages must match our observed averages (denoted by $a_k$).

$$\int g_k(x) f(x)\, dx = a_k \quad \text{for } k = 1 \ldots m$$

Solution by Calculus of Variations using Lagrange Multipliers.

# Common MaxEnt Distributions in UQ

The "shape" of the distribution is determined entirely by the information we possess.

| **Support** | **Known Constraints** | **MaxEnt Result** |
|---|---|---|
| $[a, b]$ | None (only bounds) | **Uniform** |
| $[0, \infty)$ | Mean $\mu$ | **Exponential** |
| $(-\infty, \infty)$ | Mean $\mu$, Variance $\sigma^2$ | **Gaussian** |

In UQ, if we only know the first two moments (mean and variance) and have no bounds, the Normal distribution contains the least amount of "structural" information.

# Maximum Likelihood Estimation (MLE)

When data is available $(x_1, x_2, \ldots, x_n)$, we move from subjective assignment to statistical fitting.

**The Philosophy of MLE:**

> *"Given a specific model family (e.g., Normal, Weibull), the best parameters $\theta$ are those that make the observed data most probable."*

# The Likelihood Function

Let $X$ be a random variable with PDF $f(x;\theta)$, where $\theta$ is a vector of parameters. Assuming independent and identically distributed (i.i.d.) observations, the **Likelihood function** is:

$$\mathcal{L}(\theta|x) = \prod_{i=1}^{n} f(x_i;\theta)$$

To simplify calculation (turning products into sums), we maximize the **Log-Likelihood**:

$$\ln \mathcal{L}(\theta|x) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

# Optimization

To find the estimator $\hat{\theta}_{MLE}$, we solve:

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_j} = 0 \quad \text{for all parameters } \theta_j$$

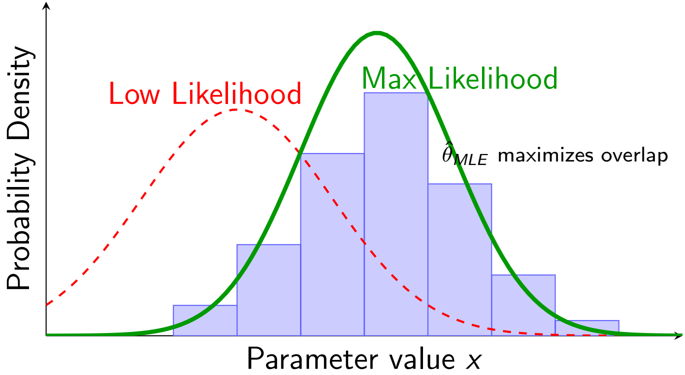
**Example: Gaussian Distribution** $\mathcal{N}(\mu, \sigma^2)$

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Maximizing $\ln \mathcal{L}$ yields the intuitive results:

$$\hat{\mu} = \frac{1}{n}\sum x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum (x_i - \hat{\mu})^2$$

# Visualizing MLE



# Model Selection

MLE gives us the best parameters **for a chosen distribution**. But which distribution family should we choose? Normal? Lognormal? Gamma?

**The Danger of Overfitting:**

- We can (almost) always increase the likelihood by adding more parameters.
- A complex model fits the *noise* in the data, not just the trend.
- In UQ, overfitting leads to poor prediction of tail risks.

# Parsimony and Information Criteria

We need a metric that rewards goodness-of-fit but penalizes complexity. In general, we want to minimize:

$$\text{Score} = -\ln(\hat{\mathcal{L}}) + \text{Penalty}(k, n)$$

Where:

- $\hat{\mathcal{L}}$ is the maximized likelihood value.
- $k$ is the number of estimated parameters.
- $n$ is the sample size.

# Akaike Information Criterion (AIC)

Derived from information theory (estimating the relative information lost).

$$\text{AIC} = 2k - 2\ln(\hat{\mathcal{L}})$$

- **Focus:** Predictive accuracy.
- **Penalty:** $2k$. It is a constant penalty per parameter.
- **Use case:** Good for complex models or when the "true" model is likely not in the candidate set. Tends to select slightly more complex models than BIC.

*Correction for small sample size ($AIC_c$) is recommended when $n/k < 40$.*

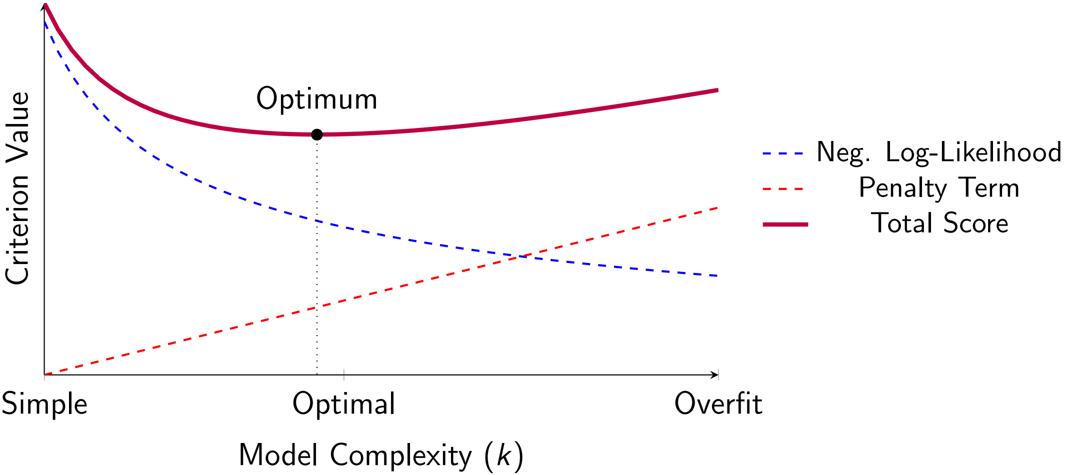
# Bayesian Information Criterion (BIC)

Derived from Bayesian probability (approximation of the Bayes Factor).

$$\mathrm{BIC} = k \ln(n) - 2 \ln(\hat{\mathcal{L}})$$

- **Focus:** Identifying the true model.
- **Penalty:** $k \ln(n)$.
- **Comparison:** Since $\ln(n) > 2$ for $n \geq 8$, BIC penalizes complexity **more heavily** than AIC.
- **Result:** BIC favors simpler models (parsimony) as datasets grow larger.

# Visualization



# Example 1

**Scenario:** We fit a dataset ($n = 100$) with two models.

**Model A (Exponential)**

- $k = 1$ parameter
- $\ln(\hat{\mathcal{L}}) = -200$

**Model B (Gamma)**

- $k = 2$ parameters
- $\ln(\hat{\mathcal{L}}) = -198$ (better fit)

**AIC Calculation:**

- $AIC_A = 2(1) - 2(-200) = 402$
- $AIC_B = 2(2) - 2(-198) = 4 + 396 = 400$

**Result:** AIC prefers Model B (lower score), despite the penalty. The gain in likelihood justified the extra parameter.

# Example 2

**Scenario:** We fit a larger dataset ($n = 1000$) with the same two models.

**Model A (Exponential)**

- $k = 1$ parameter
- $\ln(\hat{\mathcal{L}}) = -2000$

**Model B (Gamma)**

- $k = 2$ parameters
- $\ln(\hat{\mathcal{L}}) = -1998$ (better fit)

**BIC Calculation ($BIC = k \ln(n) - 2 \ln(\hat{\mathcal{L}})$):**

- $BIC_A = 1 \cdot \ln(1000) - 2(-2000) \approx 6.9 + 4000 = \mathbf{4006.9}$
- $BIC_B = 2 \cdot \ln(1000) - 2(-1998) \approx 13.8 + 3996 = 4009.8$

**Result:** BIC prefers **Model A** (lower score).

**Result:** Why? The improvement in likelihood ($+4$) was outweighed by the heavier BIC penalty for the extra parameter.

# Visual Check: Q-Q Plots

Compares your data's quantiles against the quantiles of a theoretical (usually normal) distribution.

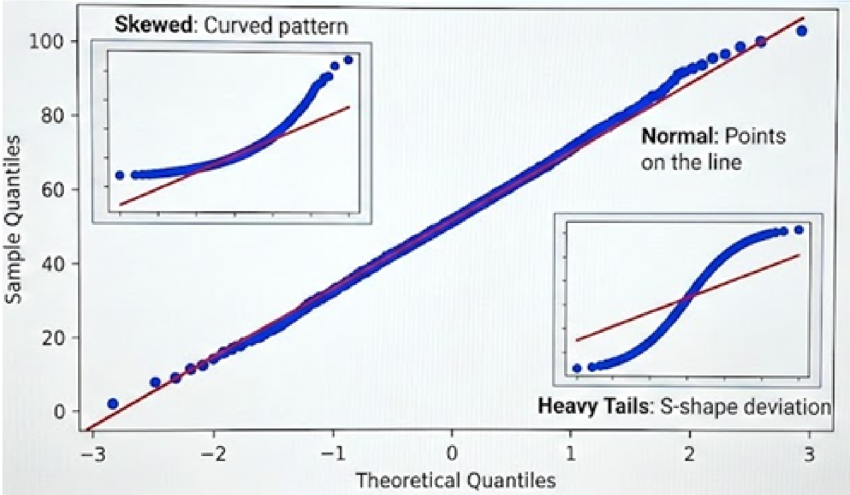
**Interpretation Rules**

- **On the Line:** Data is Normal.
- **Curved:** Data is Skewed (Left or Right).
- **S-Shaped:** Heavy Tails (Outliers).

**Why use it?**

- It is more sensitive than a histogram.
- It highlights outliers clearly.

# QQ plot



Skewed: Curved pattern
Normal: Points on the line
Heavy Tails: S-shape deviation
100
80
60
40
20
0
Sample Quantiles
−3
−2
−1
0
1
2
3
Theoretical Quantiles

# Practical Workflow

1. **Assess Data Availability:**
   - No data? $\rightarrow$ Expert Info + MaxEnt.
   - Data? $\rightarrow$ Proceed to fitting.
2. **Candidate Selection:**
   - Choose candidate PDFs based on physics (bounds, tails).
3. **Fit Models:**
   - Use MLE to find optimal parameters for each candidate.
4. **Select Best Model:**
   - Compare AIC/BIC scores.
   - Visual check (QQ-plots).

# Example: Gamma Distribution

**Definition:** A continuous distribution widely used in UQ for positive-only variables (e.g., material strength, rainfall, wind energy).

## Probability Density Function (PDF)

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0$$

**Parameters:**

- $\alpha$ **(Shape):** Controls the "peakedness" and skewness.
- $\beta$ **(Scale):** Controls the "spread" or horizontal stretching.

**Theoretical Moments:**

- Mean: $\mu = \alpha\beta$
- Variance: $\sigma^2 = \alpha\beta^2$

# Parameter Estimation Strategies

Given a sample $\mathcal{D} = \{x_1, \ldots, x_n\}$, how do we find $\hat{\alpha}$ and $\hat{\beta}$?

**1. Method of Moments (MoM)** Equate sample stats $(\bar{x}, s^2)$ to theoretical moments:

$$\hat{\beta}_{MoM} = \frac{s^2}{\bar{x}} \qquad \hat{\alpha}_{MoM} = \frac{\bar{x}^2}{s^2}$$

**2. Maximum Likelihood Estimation (MLE)** Find $\theta = \{\alpha, \beta\}$ that maximizes the likelihood $L(\theta)$:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{i=1}^{n} \ln f(x_i|\theta)$$

# UQ Softwares/Toolboxes

- UQLab
- UQ[py]Lab
- UQpy